Inaugural-Dissertation

zur Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprechts-Karls-Universität Heidelberg

Molekulare Struktur- und Funktionsanalyse der Transkriptionskontrolle der Gene *DDX3X* und *DDX3Y* in der männlichen Keimbahn

vorgelegt von
Diplombiologe
Marc-Alexander Rauschendorf
aus Tübingen

Heidelberg im Juli 2011

Die Untersuchungen zur vorliegenden Arbeit wurden von 2006 bis 2011 in der „Sektion Molekulare Genetik und Fertilitätsstörungen“ der Universitätsfrauenklinik Heidelberg unter der Leitung von Prof. Dr. HP. Vogt durchgeführt.

Gutachter der Arbeit:

Erstgutachter:	Prof. Dr. rer. nat. Hans-Peter Vogt
Zweitgutachter:	Prof. Dr. rer. nat. Herbert Steinbeisser
Tag der mündlichen Prüfung:	Heidelberg, den…….

Zusammenfassung

Molekulare Struktur- und Funktionsanalyse der Transkriptionskontrolle der Gene *DDX3X* und *DDX3Y* in der männlichen Keimbahn

Marc-Alexander Rauschendorf

Das Y-chromosomale Gen *DDX3Y* und das X-homologe Gen *DDX3X* kodieren zwei RNA-Helikasen der DEAD-Box Familie, die beide funktionell in verschiedenen Phasen der Human Spermatogenese aktiv sind (Ditton et al., 2004). Deletionen der Azoospermia Factor a (AZFa) Region in Yq11, in der *DDX3Y* lokalisiert ist, führen zu einem Totalverlust der männlichen Keimzellen, dem Sertoli Cell-only (SCO)-Syndrom. Beide Gene weisen in den kodierenden Sequenzen eine hohe Konserviertheit (92,4%) ihrer Aminosäure-Sequenzen auf, was auf eine funktionelle Selektion beider Genkopien hindeutet. Die Promoter- und 5´UTR-Sequenzen haben dagegen, seit Fehlen der Rekombination der Säuger Gonosomen, deutliche Chromosomen-spezifische Sequenzveränderungen durchlaufen. Diese Veränderungen haben zur Entstehung einer komplexen hodenspezifischen Transkriptions- und Translationskontrolle beider Gene geführt. Auf Grund der Allel-spezifischen Sequenzevolution sind unterschiedliche Core-Promotermodule zur Keimbahn-spezifischen Expressionskontrolle etabliert worden. Einige Sequenzmotive geben auch erste Hinweise auf unterschiedliche Chromatinstrukturen der beiden Promoterdomänen.

Durch die Kombination von vergleichender Genomik in sechs Säugerspezies (Mensch, Schimpanse, Rhesusaffe, Weißbüschelaffe, Rind, Maus) für beide Gene und gezielten Experimenten, konnten sowohl Allel-spezifische, als auch Spezies-spezifische *cis*-regulative Module identifiziert werden. So konnten in den Human *DDX3*(*X/Y*) Promoterregionen neun konservierte Sequenzblöcke kartiert werden. Ein solcher Sequenzblock ist der Y-spezifische *MSY2* Minisatellit (Bao et al., 2000). Eine *MSY2*, bzw. homologe *MSY2-X* Basissequenz konnte in allen Spezies für *DDX3Y* und *DDX3X* stromaufwärts zu den Transkriptionseinheiten identifiziert werden. Eine Vervielfältigung der *MSY2* Sequenz erfolgte allerdings nur in Primaten und nur in der Keimzell-spezifischen Promoterdomäne von *DDX3Y*. In den neun Human X-Y Sequenzblöcken wurden 24 X-Y konservierte Transkriptionsfaktorbindestellen (TFBS) identifiziert. Besonders auffällig ist eine in allen untersuchten Spezies X-Y konservierte SOX5-TFBS, die in den *MSY2* und *MSY2-X* Sequenzen lokalisiert ist. Insgesamt konnten 30 X-Y konservierte TFBS in den Human Promotersequenzen kartiert werden. Die Mehrzahl der dazugehörigen TFs weist eine Expression in Hodengewebe auf. Sechs gemeinsame TF-Module konnten identifiziert werden, wovon eines positionshomolog lokalisiert ist. Die konservierten TFBS und gemeinsamen TF-Module deuten somit auf einen gewissen Grad von Co-Regulation der beiden Gene in der männlichen Keimbahn hin. Daneben konnten zusätzlich mehrere Allel-spezifisch repetitive TFBS und TFBS-Cluster kartiert werden, die offensichtlich eine genspezifische Transkriptionskontrolle vermitteln. Die Kombination aus *in-silico* Analysen und gezielten Luziferase-Reportergenanalysen demonstriert somit eine erfolgreiche Strategie zur Identifizierung wesentlicher *cis*-Kontrollelemente für die *DDX3*(*X/Y*) Keimbahnexpression. Diese *cis*-Module sind daher gute Sequenzmotive für die molekulargenetische Mutationsanalyse bei infertilen Männern mit Verdacht auf Fehlfunktion der *DDX3*(*X/Y*) Expression.

Summary

Molecular structure and function analysis of the transcription control of the *DDX3X* and *DDX3Y* genes in the male germ line

Marc-Alexander Rauschendorf

The Y-chromosomal gene *DDX3Y* and ist X-chromosomal homolog *DDX3X* encode two RNA-Helikases of the DEAD-Box family. They are functional in different phases of the Human spermatogenesis (Ditton et al., 2004). Deletion of the Azoospermia Factor a (AZFa) interval in Yq11 in which *DDX3Y* is located, cause a complete loss of male germ cells. This pathology is called Sertoli Cell-only (SCO) syndrome. Both genes show in the coding sequences high conservation (92,4%) of their amino acid sequences, providing strong indication that both copies are under functional selective constraints. In contrast, the promoter and 5´UTR sequences have undergone dramatic chromosome-specific changes, probably evolved after the lack of recombination during evolution of mammalian sex chromosomes. This caused the development of a complex testis-specific transcriptional and translational control of both genes. Different core-promoter modules controlling their germ line expression arose due to the allele-specific sequence evolution. Some of the mapped sequence motives give also a lead to a probably different chromatin structure of both promoter domains.

By combination of comparative genomics of both genes in six mammalian species (human, chimpanzee, macaque, marmoset, cattle, mouse) and selective experiments, allele-specific as well as species-specific *cis*-regulative modules could be identified. By this 9 conserved sequnce blocks were mapped in the human *DDX3*(*X/Y*) promoter regions. One of these blocks is the Y-specific minisatellite *MSY2* (Bao et al., 2000). A basic *MSY2* and *MSY2-X* sequence respectively could be identified upstream of the *DDX3Y* and *DDX3X* transcription units in all analysed species. An amplification of the *MSY2* minisatellite sequence occurred only in primates and only in the germ cell specific promoter domain of *DDX3Y* gene.

24 X-Y conserved transcription factor binding sites (TFBS) could be identified in the 9 X-Y conserved sequence blocks. Most striking is one X-Y conserved SOX5-TFBS which could be localised in all species in the *MSY2* and *MSY2-X* sequences respectively. In total 30 different X-Y conserved TFBS could be mapped in both human promoter sequences. The majority of associated transcription factors show expression in testis tissue. Additionally 6 shared TF modules were identified. One of these is also positionally conserved. The conserved TFBS and shared TF modules lead to a certain degree of co-regulation of both genes in the male germ line. Aside multiple allele-specific repetitive TFBS and TFBS clusters were mapped, which mediate a gene-specific transcriptional control.

The combination of computational analysis and selective luciferase reportergene analysis demonstrated an effective strategy for identification of essential *cis*-control elements of the *DDX3*(*X/Y*) germ line expression. Hence, these *cis*-modules are prime sequence motives for molecular mutation analysis in infertile men with evidence for dysfunction of *DDX3*(*X/Y*) gene expression.

Inhaltsverzeichnis

1. Einleitung 1

1.1. DEAD-Box RNA-Helikasen 1
1.2. Die DEAD-Box RNA-Helikase DDX3X 2
1.3. Die DEAD-Box RNA-Helikase DDX3Y 5
1.4. Homologie der PI10-verwandten DEAD-Box Helikasen 7
1.5. Evolution der Säuger und Primaten 9
1.5.1. Evolution der Säugetiere 9
1.5.2. Evolution der Primaten 12
1.6. Evolution der Human Geschlechtschromosomen X und Y 13
1.7. Transkriptionskontrolle in Eukaryoten 19
1.7.1. Der Core-Promoter 19
1.7.2. Alternative Core-Promoter-Erkennungsfaktoren 25
1.7.3. Promoterstruktur, Enhancer, Mediator-Komplex und DNA-Signaturen 27
1.8. Post-transkriptionelle Regulation der Genexpression 34
1.9. Ziel der Arbeit 39

2. Material 41

2.1. Geräte und Zubehör 41
2.2. Verbrauchsmaterialien 42
2.3. Chemikalien 43
2.4. Fertigsysteme 44
2.5. Enzyme 44
2.6. Puffer und Lösungen 45
2.6.1. Isolation von genomischer DNA und Total-RNA 45
2.6.2. RNA-Gel 46
2.6.3. Gelelektrophorese 46
2.6.4. Kulturmedien und -Lösungen zur Bakterienkultivierung 47
2.7. Gewebe 47
2.8. Zellkulturen 48
2.8.1. Medien und Zusätze für Zellkulturen 48
2.8.2. Human Zellinie 49
2.8.3. Primaten Zellinien 49
2.9. Chemo-kompetente Bakterienzellen 49
2.10. Nukleotide und Nukleinsäuren 50
2.11. DNA-Größenstandards 50
2.12. Vektoren 50

2.13. Oligonukleotide für PCR und RT-PCR 51
2.14. Software und Datenbanken 51
2.14.1. Genomatix-Software zur Identifizierung von Transkriptionsfaktorbindestellen 51
2.14.2. Erstellung von Sequenzalignments mit den Programmen mVISTA, ClustalW und DiAlignTF 52
2.14.3. CpG-Island Analyse 53

3. Methoden 55

3.1. Molekularbiologische & Mikrobiologische Methoden 55
3.1.1. Isolierung genomischer DNA und Total-RNA 55
3.1.1.1. Isolierung aus Vollblut 55
3.1.1.2. Isolierung aus Kryogewebe 56
3.1.1.3. Isolierung aus PBS-Pellets 56
3.1.1.4. Kontrolle der isolierten Total-RNA 57
3.1.2. cDNA-Synthese 57
3.1.3. Amplifikation von DNA und cDNA 58
3.1.3.1. Standard Polymerase-Kettenreaktion 58
3.1.3.2. Schnittstellengenerierung mit Hilfe modifizierter Oligonukleotide 60
3.1.3.3. Mutationsgenerierung mit Hilfe modifizierter Oligonukleotide 60
3.1.4. Klonierungstechniken 61
3.1.4.1. Restriktionsspaltung doppelsträngiger DNA 61
3.1.4.2. Dephosphorylierung von Vektoren 62
3.1.4.3. Ligation 62
3.1.4.4. Transformation kompetenter *E. coli*-Zellen 63
3.1.4.5. Präparation von Plasmid-DNA aus Bakterien 64
3.1.4.6. Lagerung von Bakterienkulturen 64
3.1.5. Ethanolpräzipitation von Nukleinsäuren 64
3.1.6. Konzentrationsbestimmung von Nukleinsäuren durch Messung der Optischen Dichte 65
3.1.7. DNA-Sequenzierung – nach Kettenabbruchverfahren 65
3.1.7.1. Ansatz der Sequenzreaktion 67
3.1.8. Agarose-Gelelektrophorese 67
3.1.9. Photodokumentation der Agarosegele 68
3.1.10. DNA-Extraktion aus Agarosegelen 68
3.1.11. *In-vitro* Transkription/Translation von *DDX3X* 5´UTRs 68
3.2. Zellbiologische Methoden 72
3.2.1. Kultivierung und Lagerung von humanen Zelllinien 72
3.2.2. Promotor- und Enhanceranalysen im zellulären System 73
3.2.3. Transiente Transfektion von Reportergenkonstrukten in eukaryotischen Zellen mittels Transfektion mit nicht-liposomalen Lipiden 73
3.2.4. „Dual-Luciferase Reporter Assay“ Methode 74

4. Ergebnisse 77

4.1. Keimbahn-Komplexität der *DDX3Y* Transkripte und Kartierung der dazugehörigen Core-Promotermotive 77
4.1.1. Kartierung der *DDX3Y* Transkriptionsstartstelle I (TSS-I) 78
4.1.2. Kartierung der potentiellen *DDX3Y* TSS Region im *MSY2* Minisatellitenrepeat 83
4.1.3. Kartierung weiterer potentieller TSS-Regionen im *DDX3Y* Exon-T 89
4.2. Keimbahn-Komplexität der *DDX3X* Transkripte und Kartierung der dazugehörigen Core-Promotermotive 92
4.2.1. Kartierung der *DDX3X* Transkriptionsstartstelle I (TSS-I) 97
4.2.2. Kartierung der selektiv genutzten *DDX3X* Initiationsstelle (T/B-TSS-II) im TSS-Bereich II 98
4.2.3. Kartierung der hodenspezifisch genutzten *DDX3X* Transkriptionsstartstelle III (T-TSS-III) 100
4.3. Vergleich der kartierten *DDX3Y* und *DDX3X* TSS-Regionen 102
4.4. Evolution der *DDX3Y* Keimbahn Core-Promoterstruktur in verschiedenen Säugerspezies 105
4.4.1. Evolution der *DDX3Y* Keimbahn Core-Promoterstruktur in Primaten 105
4.4.2. Evolution der *DDX3Y*/*Ddx3y* Keimbahn Core-Promoterstruktur in Rind und Maus 120
4.5. Evolution der *DDX3X* Keimbahn Core-Promoterstruktur in verschiedenen Säugerspezies 129
4.5.1. Evolution der *DDX3X* Keimbahn Core-Promoterstruktur in Primaten 129
4.5.2. Experimentelle Überprüfung eines computergenerierten *M. mulatta DDX3X* Transkriptes 138
4.5.3. Evolution der *DDX3X*/*Ddx3x* Keimbahn Core-Promoterstruktur in Rind und Maus 139
4.6. Kartierung der potentiell funktionellen Promoterdomänen von Human *DDX3Y* und *DDX3X* 146
4.6.1. VISTA-Plot und Analyse konservierter Bereiche in der Human *DDX3Y* und *DDX3X* Promoterregion 147
4.6.2. Analyse der *DDX3Y* und *DDX3X* Promoterregionen auf DNA-Signaturen mit binärer und homopolymerer Nukleotidkomposition 152
4.6.3.1. Analyse konservierter Transkriptionsfaktorbindestellen und TFBS-Mustererkennung in den Human *DDX3Y* und *DDX3X* Promoterregionen 160
4.6.3.2. TFBS-Mustererkennung in den Human *DDX3Y* und *DDX3X* Promoterregionen 169
4.6.3.3. TF-Module in den Human *DDX3Y* und *DDX3X* Promoterregionen 176
4.6.3.4. Konservierte Abstände zwischen cTFBS 179
4.7. Vertikaler Vergleich der Human *DDX3Y* und *DDX3X* Promoterdomänen mit homologen Sequenzen anderer Säugetierspezies 181
4.8. Unterschiedliche Gonosomenevolution im Bereich der *DDX3Y* und *DDX3X* Loci 202
4.8.1. *DDX3Y* und *DDX3X* Exon- und Intronstrukturen 202
4.8.2. S/MAR-Sequenzen und *Alu*-Elemente 205
4.8.3. *DDX3X* Intron1 Gen 211
4.9. Funktionelle Analysen zur Human *DDX3X* Promoterregion 213
4.9.1. Funktionelle Kartierung der *DDX3X* Promoterdomäne 213
4.9.2. Kartierung partieller Funktionsbereiche im *DDX3X* Promoter 219
4.10. Funktionelle Analysen zum *DDX3Y MSY2* Minisatellitenrepeat und zur *DDX3X MSY2-X* Sequenz 228

4.10.1. Funktionelle Analyse des *DDX3Y MSY2* Minisatellitenrepeats 231
4.10.2. Funktionelle Mutationsanalyse des *DDX3Y MSY2* Minisatellitenrepeats 235
4.10.3. Funktionelle Analyse der *DDX3X MSY2-X* Sequenz 238
4.10.4. Funktionelle Mutationsanalyse der *DDX3X MSY2-X* Sequenz 241
4.11. Analyse potentieller Translationskontroll-Elemente in der 5´UTR Sequenz der *DDX3X* Transkripte mit möglichem Einfluß auf die Translation in humanem Hodengewebe 243
4.11.1. *DDX3X* 5´UTR *cis*-Elemente mit potentieller Funktion für die Translationskontrolle in Hodengewebe 244
4.11.2. Strukturelle *cis*-Elemente in der *DDX3X* und *DDX3Y* 5´UTR-Sequenz mit potentieller Translationskontrollfunktion 249
4.11.3. *In-silico* Vorhersage der optimalen *DDX3X* und *DDX3Y* RNA-Sekundärstruktur nach Zuker-Algorithmus 253
4.11.4. Vertikaler Vergleich potentiell cis-regulativer Elemente der *DDX3X* und *DDX3Y* 5´UTR-Sequenzen 260
4.11.5. *In-vitro* Translations-Assay zur Messung der Translationseffizienz unterschiedlicher Human *DDX3X* Transkriptvarianten 266

5. Diskussion 276

5.1. Konserviertheit und spezifische Evolution der *DDX3X* und *DDX3Y* Core-Promotermotive innerhalb der Säuger 277
5.1.1. *DDX3X* Transkriptionsstartstellen 279
5.1.2. *DDX3Y* Transkriptionsstartstellen 283
5.2. Auffällige Nukleotidkompositionen und „Simple Sequence Repeats“ im Bereich der *DDX3X* und *DDX3Y* Core-Promoterregionen 287
5.2.1. Nukleotidkompositionen im Bereich der *DDX3X* Core-Promotermodule 287
5.2.2. Nukleotidkompositionen im Bereich der *DDX3Y* Core-Promotermodule 288
5.3. Konserviertheit und spezifische Evolution der *DDX3X* und *DDX3Y* Promoterregion 289
5.3.1. Konservierte TFBS und TF-Module 291
5.3.2. CpG-Islands 294
5.4. Potentiell funktionelle Chromatinmodule für die hodenspezifische Transkriptionsaktivierung der *DDX3X* und *DDX3Y* Genloci 296
5.4.1. *DDX3X* Genexpressionskontrolle durch ZF5 Homomer-Chromatinkomplexe 296
5.4.2. *DDX3Y* Genexpressionskontrolle durch S/MARs 297
5.4.3. S/MARs und deren Assoziation zur Kernmatrix 299
5.4.4. S/MARs und *DDX3*(*X/Y*) 3´UTR-Bildung 300
5.4.5. *Alu*-Sequenzen und *DDX3X* 3´UTR-Bildung 301
5.5. Funktionelle *DDX3X* Promoteranalysen 302
5.5.1. Kartierung cis-regulativer Module in der *DDX3Y* Promoterregion 302
5.5.2. Potentieller Einfluß von G-Quadruplex-Sequenzen auf die *DDX3X* Promoteraktivität 304
5.5.3. Weiterführende Projekte zur Bestimmung der *DDX3X* Promoterfunktionen 306

5.6. Funktionelle *MSY2* und *MSY2-X* Analysen 307
5.7. Funktionelle *DDX3X* 5´UTR Analyse 308
5.7.1. *DDX3X* 5´UTR und 5´UTR-Splicing 309
5.7.2. *DDX3X* 5´UTR und uORFs 309
5.7.3. *DDX3X* 5´UTR und der potentielle Einfluß auffälliger Sequenzmotive 311
5.7.4. Weiterführende Projekte zur Bestimmung der *DDX3X* 5´UTR Funktionen 312
5.8. Ausblick 315
5.8.1. Diagnostische Möglichkeiten der *DDX3Y* Analyse 316

6. Literaturverzeichnis 318

7. Abbildungsverzeichnis 360

8. Tabellenverzeichnis 366

9. Abkürzungsverzeichnis 372

10. Anhang 376

11. Veröffentlichungen 407

12. Eidesstattliche Erklärung 409

13. Danksagung 411

1. Einleitung

1.1. DEAD-Box RNA-Helikasen

Die Human RNA-Helikasen DDX3X und DDX3Y gehören zur Familie der DEAD-Box Helikasen, die zur Helikasen-Superfamilie 2 (SF2) gezählt werden (Cordin et al., 2006). Das namensgebende DEAD-Motiv (Asp-Glu-Ala-Asp) und acht weitere Motive sind in allen Proteinen dieser Familie, von den Bakterien bis zum Menschen, hochkonserviert (Abb. 1-1; Linder et al., 1989; Rocak & Linder, 2004; Cordin et al., 2006; Linder, 2006).

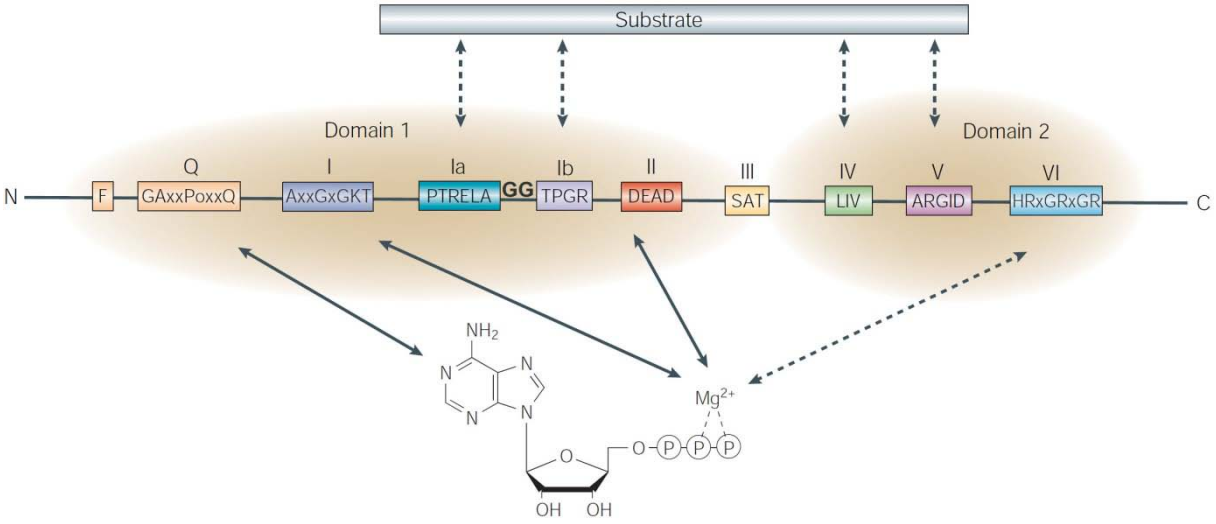


Abb. 1-1: Schematische Darstellung der konservierten Motive in DEAD-Box RNA-Helikasen.
Dargestellt sind die neun hochkonservierten Proteinmotive (Q bis VI) der DEAD-Box Helikasen. Stromaufwärts vor dem Q-Motiv ist ein hochkonserviertes Phenylalanin (F) mit dem Q-Motiv assoziiert. In vielen DEAD-Box Helikasen ist zwischen den Motiven Ia und Ib eine Glycin-Dublette (GG) zu finden. Die neun Motive sind in zwei funktionelle Proteindomänen 1 und 2 unterteilt. Die Motive Ia, Ib, IV und V sind an der RNA-Bindung beteiligt. Das Q-Motiv ist u.a. für die ATP-Bindung verantwortlich. Motiv I und II, auch Walker A und Walker B nach dem Erstbeschreiber genannt, sind für die ATPase- und Helikase-Aktivität notwendig. Das Motiv III scheint je nach Konformation in Domäne 1 und in der Furche zwischen 1 und 2 lokalisiert zu sein. Für Motiv III wird eine Verknüpfung von ATPase- und Helikase-Funktion angenommen. Intramolekulare Interaktionen zwischen den einzelnen Motiven sind für Motiv I, Ia, II, V und VI bekannt. Großbuchstaben geben die hochkonservierten Aminosäuren im Einbuchstabencode an. Der Kleinbuchstabe x steht für jegliche Aminosäure, die an entsprechender Position vorkommen kann. Der Kleinbuchstabe o im Q-Motiv steht für die Aminosäuren Serin und Threonin mit Hydroxylgruppe. Graphik aus Rocak & Linder (2004) verändert dargestellt.

Die neun Motive sind zentral im Protein lokalisiert und erstrecken sich über ~ 350 Aminosäuren (AS). Außerhalb dieser konservierten Region weisen DEAD-Box RNA-Helikasen im N- und C-Terminus zumeist nur geringe Homologien auf. DEAD-Box RNA-Helikasen werden funktionell für jeden Schritt des RNA-Metabolismus beschrieben (Cordin et al., 2006; Rosner & Rinkevitch, 2007; Schröder, 2010). Angefangen bei der Transkriptionsregulation als Co-Faktor für Transkriptionsfaktoren (TFs), über RNA-Splicing, RNA-Export, Ribosomen-Biogenese, Translationsinitiation bis zum RNA-Abbau sind DEAD-Box Proteine beteiligt.

Die Vielfalt der unterschiedlichen Prozesse wird über die nicht konservierten N- und C-terminalen Regionen der DEAD-Box Proteine realisiert (Kwong et al., 2005). Als eine der zentralen und phylogenetisch konservierten Funktionen der DEAD-Box RNA-Helikasen gilt das Entwinden von RNA-Strukturen, in großen mRNP-Komplexen („messenger ribonucleoprotein complexes“; Jankowsky & Putnam, 2010; Pan & Russel, 2010). Diese als RNA-Chaperon beschriebene Funktion ermöglicht das lokale Entwinden von RNA-Sekundärstrukturen ohne die gesamte RNA-Struktur zu beeinflussen (Yang & Jankowsky, 2005; Liu et al., 2008; Del Campo et al., 2009). Das Entwinden und Remodellieren der RNA-Strukturen geschieht unter ATP-Verbrauch, wobei offensichtlich die Energie aus der Hydrolyse nicht für die Helikaseaktivität selbst, sondern für die Reaktivierung des Enzyms notwendig ist (Liu et al., 2008). Da DEAD-Box Proteine ihre RNA-Substrate generell nicht sequenzspezifisch selektionieren, ist bisher noch nicht geklärt, wie die Substratauswahl geschieht. Diskutiert werden eine RNA-Substratauswahl im Verbund des mRNP-Komplexes und über spezifische RNA-Strukturen, anstatt durch Erkennung spezifischer mRNA-Sequenzen (Bowers et al., 2006; Jankowsky & Bowers, 2006; Lai et al., 2010).

1.2. Die DEAD-Box RNA-Helikase DDX3X

Der *DDX3X* Genlocus, auch *DBX* („DEAD box protein 3, X-chromosomal“) genannt, ist auf dem kurzen Arm des Human X-Chromosoms in Xp11.3→ p11.23 lokalisiert (Chung et al., 1995; Park et al, 1998). Die *DDX3X* Genstruktur besteht aus 17 Exonen, die für ein DEAD-Box Protein von 662 AS und ~73kDa[1] kodieren (Kim et al., 2001). Zusätzlich zu diesem funktionellen Genlocus ist ein intronloser *DDX3* Pseudogen-Locus auf den Human Chromosom 4q24 beschrieben (Kim et al., 2001). Ein zweiter intronloser *DDX3* Pseudogen-Locus wurde in Xq13 identifiziert (Kim et al., 2001). Neuere Untersuchungen haben ergeben, daß es sich um ein *DDX3Y* Pseudogen handelt (Chang & Liu, 2010). *DDX3X* entgeht der X-Inaktivierung und hat ein homologes Gen, *DDX3Y*, im proximal Bereich Yq11 des Human Y-Chromosoms (Lahn & Page, 1997). Die Human *DDX3X* Expression wird für RNA, sowie Protein als ubiquitär beschrieben (Lahn & Page, 1997; Kim et al., 2001). Für die adulte männliche Keimbahn konnte eine postmeiotische DDX3X Proteinexpression in Spermatiden gezeigt werden (Abb. 1-2; Ditton et al., 2004).

Das DDX3X Protein hat, in Assoziation mit verschiedenen mRNP-Komplexen, sowohl nukleäre als auch zytoplasmatische Funktionen und kann offensichtlich zwischen beiden Zell-kompartimenten pendeln (Schröder, 2010). Im Zellkern ist DDX3X an der Transkriptions-

[1] *DDX3X* mRNA GenBank Acc.No.NM_001356; DDX3X Protein GenBank Acc.No. NP_001347

kontrolle der Gene p21^{waf}, E-Cadherin und INF-β (Interferon beta) z.B. als Co-Faktor des TFs SP1 oder im Verbund mit IRF3 („Interferon regulatory factor 3“) beteiligt (Chao et al., 2006; Schröder, 2010). Soulat et al. (2008) konnten zeigen, daß die IFN-Promoter-aktivierung durch DDX3X in Abhängigkeit der posttranslationalen DDX3X-Phosphorylierung, durch die „TANK-binding Kinase 1“ (TBK1), steht. Diese für den NF-κB Signalübertragungsweg wichtige Serin/Threonin-Proteinkinase TBK1 wird ubiquitär mit höchster Expression in Hodengewebe transkribiert (Tojima et al., 2000).

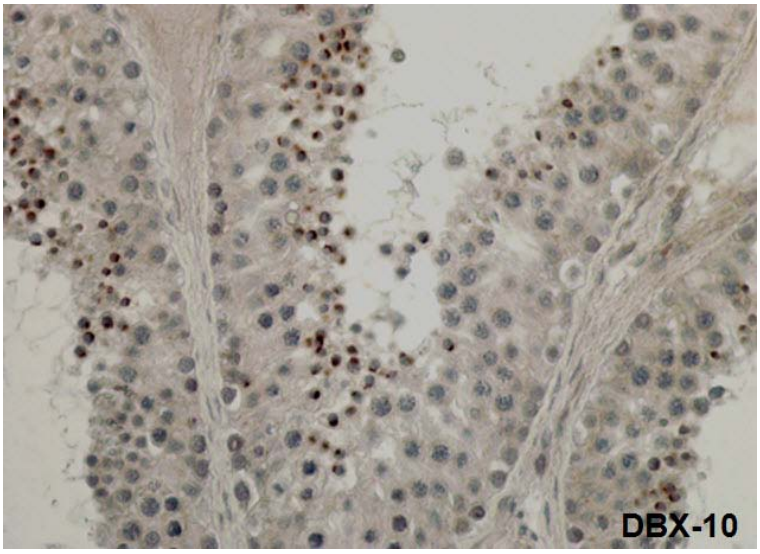


Abb. 1-2: Lokalisation des DDX3X Proteins in einem Human Hodengewebsschnitt.
DDX3X wird in den gezeigten Hodentubuli ausschließlich in Spermatiden exprimiert (braune Färbung). DAB-Färbung („3,3´-Diaminobenzidin Tetrachlorid“) nach Inkubation mit DDX3X-spezifischem Antikörper DBX-10. Gegenfärbung mit Hämatoxylin II. (Nach Ditton et al., 2004).

Ebenfalls im Zellkern scheint DDX3X mit dem Spliceosom und gesplicten mRNAs zu interagieren (Merz et al., 2007, Schröder, 2010). Eine direkte Beteiligung am Splicingprozeß konnte für DDX3X bisher nicht gezeigt werden. Die DDX3X Assoziation mit mRNPs des nukleären Exportmechanismus, wie NXF1/TAP („Nuclear RNA export factor 1“ oder „Tip-associated Protein“) und CRM-1 (Exportin-1) führen zur Annahme, daß DDX3X selbst nicht aktiv an Splicing- und Exportmechanismen beteiligt ist, sondern daß es die mRNAs vom Zellkern ins Zytoplasma als Co-Faktor zur Translation begleitet (Yedavalli, et al., 2004; Lai et al., 2008; Schröder, 2010). Für einen Großteil der nukleären mRNAs in Eukaryoten ist der NXF1/TAP-vermittelte mRNA-Export der Standard-Exportmechanismus (Stutz & Izaurralde, 2003; Lévesque et al., 2006). Eine generelle Beteiligung von DDX3X am mRNA-Export scheint jedoch nicht gegeben, so daß dieser Mechanismus nur für bestimmte mRNA-Populationen gilt.

Die zytoplasmatische DDX3X Funktion scheint v.a. in der Entwindung komplexer 5´UTR-Strukturen (untranslated region) während der Translation zu bestehen und somit ein schnelles und effizientes Scannen des Translationsapparates entlang strukturierter, langer 5´UTRs zu

gewährleisten (Tarn & Chang, 2009; Schröder, 2010). Die DDX3X Assoziation mit dem eIF3 Multiprotein-Komplex („eukaryotic translation initiation factor 3“), der mit der 40S Ribosomen-Untereinheit und dem eIF4F Multiprotein-Komplex verbunden ist, sowie Beobachtungen aus DDX3X-Knockdown und DDX3X-Mutationsexperimenten bestätigen die Annahme, daß DDX3X für die Translation strukturierter 5´UTRs, nicht aber generell für die mRNA-Translation, benötigt wird (Yedavalli et al., 2004; Lai et al., 2008; Lee et al., 2008). Zusätzlich zu DDX3X scheinen die Proteinlevel von eIF4E, dem 5´Cap-Bindeprotein, sowie eIF4A, der generellen DEAD-Box RNA-Helikase des Translationsinitiationskomplexes eIF4F, kritisch für die Translationsinitiation langer, komplexer 5´UTRs zu sein (von der Haar et al., 2004).
In einer Arbeit konnte eine DDX3X Funktion als Repressor der Translationsinitiation gezeigt werden, indem DDX3X die Bindung von eIF4E an die mRNA 5´Cap-Struktur inhibiert (Shih et al., 2008). In diesem Fall wirkt DDX3X als eIF4E-BP (Bindeprotein) und verhindert eine normale 5´Cap-abhängige Translationsinitiation (von der Haar et al., 2004; Richter & Sonnenberg, 2005; Jackson et al., 2010). Durch die DDX3X-spezifische Repression der 5´Cap-abhängigen Translationsinitiation sehen Shih et al. (2008) eine Möglichkeit die alternative IRES-abhängige Translationsinitiation („Internal Ribosome Entry Site“) zu fördern. Da es sich bei diesem Ergebnis aber um ein Zellkulturexperiment mit überexprimiertem DDX3X Protein handelt, ist es auch möglich, daß die Höhe des DDX3X-Proteinlevels einen zusätzlichen Einfluß auf die DDX3X Translationsfunktion nimmt.

Weiterhin ist DDX3X in die Zellzykluskontrolle und somit das Zellwachstum involviert (Fukumura et al., 2003; Sekiguchi et al., 2004; Lai et al., 2010). Die temperatur-sensitiv (ts) letale Hamster-zellinie tsET24, ein Derivat der BHK21 Zellinie, trägt eine Punktmutation im Hamster DDX3X Protein (Mutation P267S zwischen Motiv I und Ia). Sie führt zu einem Zellzyklus-Arrest in der G1-Phase (Fukumura et al., 2003). Sowohl die Expression von Human DDX3X-Protein, als auch von Human DDX3Y, sind in der Lage, die Hamster DDX3X-Mutante funktionell zu ersetzen und eine Transition des Zellzykluses von der G1- zur S-Phase zu ermöglichen (Fukumura et al., 2003; Sekiguchi et al., 2004). Lai et al. (2010) konnten zeigen, daß der zugrundeliegende Mechanismus der DDX3X Zellzyklus-Funktion, in der Translationskontrolle der Cyclin-E1 mRNA liegt. Der Zellzyklus-Regulator Cyclin-E1 wird für die Aktivierung der CDK2 („cyclin-dependent kinase 2“, oder „Cell division protein kinase 2“) in der späten G1-Phase benötigt (Möröy & Geisen, 2004). Als Cyclin-E1/CDK2-Komplex sind diese Faktoren für die Expression spezifischer Gene der S-Phase verantwortlich. Lai et al. (2010) zeigen in mehreren Knockdown- und Reportergen-Experimenten, daß DDX3X für die Translation der Cyclin-E1 mRNA benötigt wird. Die 5´UTR von Cyclin-E1 ist sehr (G,C)-reich und verfügt über das Potential zur stabilen Sekundärstrukturbildung (Lai et al., 2010). Die oben beschriebene DDX3X Funktion, als unterstützende RNA-Helikase bei der Translation komplexer 5´UTR-Strukturen, ist

für die Cyclin-E1 mRNA-Translation notwendig und macht DDX3X somit zu einem Zellzyklus-Regulator. Beim Übergang von der G2-Phase in die Mitose scheint DDX3X selbst Ziel post-translationaler Kontrolle zu werden, indem das Threonin T204 im Q-Motiv phosphoryliert wird (Sekiguchi et al., 2007). Die Phosphorylierung durch den Cyclin-B/CDK1-Komplex führt zur Inaktivierung des DDX3X Proteins. Der Cyclin-B/CDK1-Komplex ist für die Transition der G2-Phase in die Mitose essentiell (Porter & Donoghue, 2003; Miyazaki & Arai, 2007; Gavet & Pines, 2010; Lindqvist, 2010). Insgesamt scheinen viele der hier aufgeführten DDX3X Funktionen evolutionär von Hefe bis Mensch konserviert zu sein (Chuang et al., 1997; Tarn & Chang, 2009).

Interessanterweise ist DDX3X das Ziel mehrerer Viren, die die DDX3X-Proteinfunktionen für ihre eigenen Zwecke manipulieren (Schröder, 2010). Auf Grund der limitierten Virengenome greifen die Viren Schlüsselfaktoren der Wirtszelle an, um sich selbst replizieren zu können. HBV (Hepatitis-B-Virus), HCV (Hepatitis-C-Virus), HIV (Humanes Immundefizienz-Virus) und VACV (Vaccinia Virus aus der Familie der Pockenviren) können z.B. die DDX3X/CRM-1 Kern-exportfunktion ausnutzen, um virale RNAs aus dem Kern ins Zytoplasma zu schleusen, oder um die antivirale Immunantwort über IFN-β zu inhibieren (Schröder, 2010). Im Fall von DDX3X wenden die Viren also mehrere Strategien an, um einerseits die Replikation des Virengenoms zu sichern und andererseits um die Immunantwort der Wirtszelle zu unterdrücken (Schröder, 2010). DDX3X konnte ebenfalls in einem antiapoptotischen Proteinkomplex, bestehend aus GSK3 („Glykogen Synthase-Kinase-3“), DDX3X, cIAP-1 („cellular Inhibitor of Apoptosisprotein-1“), identifiziert werden, der eine induzierte Apoptose über einen Todesrezeptor verhindert (Sun et al., 2008). Eine weitere DDX3X Beteiligung in der Tumorprogression scheint über die Überregulation des TFs SNAI1 (Protein snail Homolog) zu bestehen, der in Metastasenbildung und Zellmigration, EMT genannt („epithelial to mesenchymal transition“), beteiligt ist (Sun et al., 2011).

1.3. Die DEAD-Box RNA-Helikase DDX3Y

Der *DDX3Y* Genlocus, auch *DBY* („DEAD box protein 3, Y-chromosomal“) genannt, ist auf dem langen Arm des Human Y-Chromosoms in proximal Yq11.21 lokalisiert (Abb. 1-4; Vogt et al., 1996; Lahn & Page, 1997). Die *DDX3Y* Genstruktur besteht aus 17 Exonen, die für ein DEAD-Box Protein von 660 AS und ~73kDa[2] kodieren. Die Aminosäuresequenzen der homologen Proteine DDX3Y und DDX3X weisen eine AS-Sequenzidentität von 92,4% auf. Die Human *DDX3Y* mRNA-Expression gilt als ubiquitär (Lahn & Page, 1997), aber die DDX3Y

[2] *DDX3Y* mRNA GenBank Acc.No. NM_004660; DDX3Y Protein GenBank Acc.No. NP_004651

Proteinexpression ist auf prämeiotische, männliche Keimzellen (Spermatogonien und präleptotäne Spermatozyten) in den adulten Hodentubuli beschränkt (Abb. 1-3; Ditton et al., 2004).

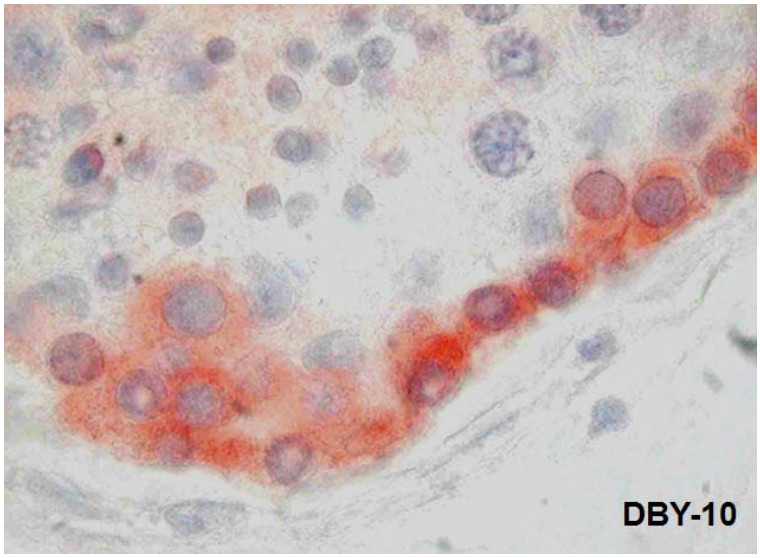


Abb. 1-3: Lokalisation des DDX3Y Proteins in einem Human Hodengewebsschnitt.
DDX3Y wird in dem gezeigten Hodentubulus in Spermatogonien exprimiert (rote Färbung). ACE-Färbung („Acetylcarbazol“) nach Inkubation mit DDX3Y-spezifischem Antikörper DBY-10. Gegenfärbung mit Hämatoxylin II. (Nach Ditton et al., 2004).

Der *DDX3Y* Genlocus liegt im AZFa-Deletionsintervall („Azoospermia Faktor a“; Vogt et al., 1996), der dem zuvor beschriebenen Deletionssubintervall 5C entspricht (Vollrath et al., 1992). Männer mit SCO-Syndrom (Sertoli Cell-only; OMIM ID # 400042) zeigen gehäuft eine 792kb lange, de novo auftretende Y-chromosomale Mikrodeletion der AZFa-Region (Abb. 1-4), die zu einem totalen Verlust der männlichen Keimzellen führt (Vogt, 2005). Als Ursache der AZFa-Deletion wurden zwei homologe HERV15-Sequenzblöcke („Human Endogenous Retrovirus“), HERV15yq1/q2, identifiziert, die die AZFa-Region flankieren (Kamp et al., 2000). Solche intrachromosomalen Rekombinationsereignisse, auch NAHR genannt („Non-allelic Homologous Recombination“), führen, v.a. während der Gametogenese, zu chromosomalen Rearrangements (McVean, 2010; Sasaki et al., 2010).

Da das DDX3Y Protein ausschließlich in männlichen Keimzellen exprimiert wird, erleiden Patienten mit AZFa-Deletion keine somatischen Schäden (Ditton et al., 2004; Vogt et al., 2008). Im AZFa Deletionsintervall sind zwei funktionelle Gene *USP9Y* („Ubiquitin-specific protease 9, Y chromosome“) und *DDX3Y* lokalisiert (Abb. 1-4). Zum AZFa-Locus wird zusätzlich das Gen *UTY* gezählt, welches distal des *DDX3Y* Gens lokalisiert ist und nicht in der AZFa-Deletion liegt (Vogt et al., 2007). Isolierte AZFa Gendeletionen oder -Mutationen konnten bisher nur für das *USP9Y* Gen beschrieben werden. Im Gegensatz zur vollständigen AZFa Deletion rufen *USP9Y* Mutationen keinen Totalverlust der männlichen Keimzellen hervor, sondern führen zu Phänotypen mit reduzierter Fertilität, die familiär erblich sein können (Krausz et al., 2006; Luddi et al., 2009). *USP9Y* Genanalysen in Primaten zeigten zudem, daß *USP9Y* Mutationen in diesen Spezies nicht zu wesentlichen Fertilitätsverlusten führen (Tyler-Smith, 2008; Goto et al.,

2009). Isolierte *DDX3Y* wie *DDX3X* Deletionen konnten bisher nicht dokumentiert werden, was für eine frühe Letalität dieser Mutationen in Embryonen sprechen könnte.

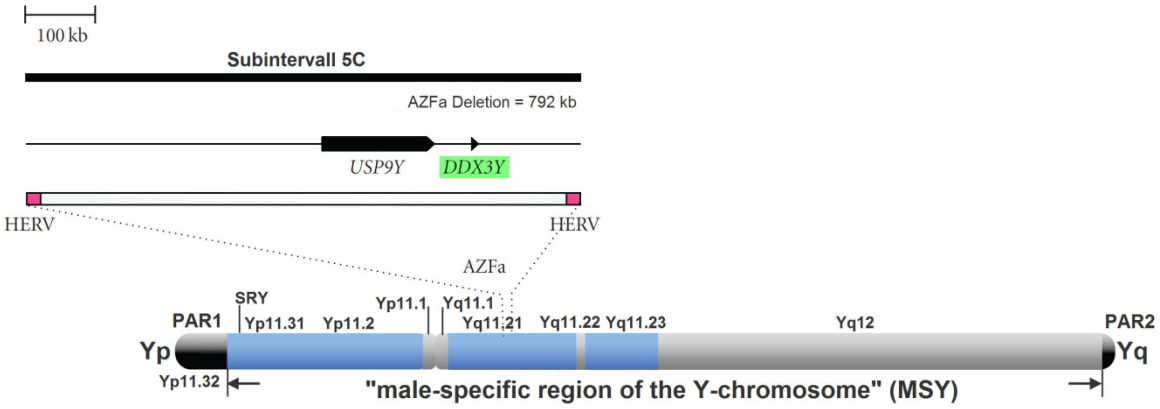


Abb. 1-4: Schematische Darstellung des Human Y-Chromosoms und der AZFa Deletion.
Das zentrale Idiogramm stellt das humane Y-Chromosom dar. Der kurze Arm des Y-Chromosoms wird als Yp11, die euchromatischen Bereiche des langen Armes werden als Yq11 und die heterochromatischen Abschnitte des langen Armes werden als Yq12 bezeichnet. Das Euchromatin ist im Idiogramm in blau, das Heterochromatin ist in grau dargestellt. Die „Pseudoautosomalen Regionen" PAR1 und -2 an den äußersten Enden sind in schwarz eingezeichnet. Alles zwischen den terminalen PARs wird als MSY („male-specific region of the Y-chromosome") bezeichnet. Oberhalb des Idiogramms ist die genomische Organisation der AZFa Region dargestellt, die aus den zwei „single-copy Genen" *USP9Y* und *DDX3Y* (in 5´→3´Polarität) besteht. *DDX3Y* ist grün hervorgehoben. Die AZFa Region wird von zwei homologen HERV15yq1/q2 („Human Endogenous Retrovirus #15") Sequenzblöcken flankiert, die für die AZFa Deletion, durch NAHR („Non-allelic Homologous Recombination") verantwortlich sind. Graphik verändert dargestellt aus Navarro-Costa et al., 2010; Skaletsky et al, 2003; Foresta et al., 2001; Lahn & Page, 1997; Vogt et al., 1996; Vollrath et al., 1992.

1.4. Homologie der PL10-verwandten DEAD-Box Helikasen

Innerhalb der DEAD-Box Helikasen weisen die Subfamilien DDX3/Ded1p[3] und die DDX4/VASA ebenfalls eine hohe Homologie in den sonst variablen N- und C-Termini der Aminosäuresequenzen auf (Mochizuki et al., 2001; Johnstone et al., 2005; Rosner & Rinkevich, 2007). Interessanterweise ist in diesen zwei Subfamilien mehrfach eine keimbahnspezifische Funktion entstanden. Die DDX4-Subfamilie ist wohl durch Duplikation des Ur-DDX3-Gens, als *PL10* bezeichnet, entstanden und hat eine reine Keimbahnfunktion angenommen (Leroy et al., 1987, 1989; Mochizuki et al., 2001). Sie ist nur in den Metazoa, nicht aber in Hefen, Pilzen und Pflanzen anzutreffen. Innerhalb der Säugetiere ist mit DDX3Y dann ein weiteres PL10-Homolog mit spezifischer Keimbahnfunktion in *Homo sapiens* entstanden. Spezifisch für die Maus konnte zudem ein funktionelles, autosomales *Ddx3x*-Retrogen, *D1Pas1* genannt, identifiziert werden (Session et al., 2001). D1Pas1 wird spezifisch in männlichen Keimzellen exprimiert und substituiert das wahrscheinlich nicht funktionelle Maus Ddx3y Protein (Mazeyrat et al., 1998; Session et al., 2001; Vong et al., 2006). Weitere funktionelle *DDX3*-Retrogene sind bisher nicht

[3] Funktionelle *DDX3*-Gene: Human: *DDX3X* & *DDX3Y*, Maus: Ddx3x & *D1Pas1*/PL10, *Xenopus*: *an3*, *Drosophila*: *bel*/belle, Hefe: *DED1*/Ded1p.

bekannt, obwohl in mehreren Säugerspezies, wie z.B. in Schimpanse, Orang-Utan und Rind prozessierte *DDX3*-Pseudogene identifiziert wurden (Liu et al., 2009; Chang & Liu, 2010). Das autosomale Rind *PL10*-Retrogen ist zwar nicht proteinkodierend, aber es konnte gezeigt werden, daß es transkriptionell aktiv und möglicherweise auch regulativ aktiv ist (Chang & Liu, 2010). Es ist somit nicht auszuschließen, daß transkribierte *DDX3*-Pseudogene regulativ, als antisense RNAs, wirken können (Balakirev & Ayala, 2003; Korneev et al., 2005; Khachane & Harrison, 2009; Alexander et al., 2010).

Die Entstehung des autosomalen *DDX3*-Retrogens *D1Pas1* ermöglicht der Maus das DDX3-Protein während der Meiose in den Spermatozyten zur Verfügung zu stellen. In den pachytänen Spermatozyten (Prophase I) der Säugetiere werden in männlichen Keimzellen das X- und Y-Chromosom transkriptionell inaktiviert. Die gepaarten Gonosomen werden in einer bestimmten Kerndomäne als „XY-Body“ sichtbar (Turner, 2007). Dieses chromosomenweite Chromatin-remodelling wird als „meiotic sex chromosome inactivation“ (MSCI). Die Entstehung von Retrogenen gonosomenlokalisierter Gene hat in der Säugerentwicklung mehrfach stattgefunden und ermöglicht so eine spezifische Proteinexpression in Spermatozyten, da die Autosomen nicht von der Inaktivierung betroffen sind (Wang, 2004). Nach der Meiose wird das XY-Body aufgelöst. Die Gonosomen verbleiben in Spermatiden allerdings transkriptionell relativ inaktiv (Namekawa et al., 2006; Turner, 2007). Die reprimierenden Eigenschaften des MSCI-Chromatins bleiben erhalten und werden als „postmeiotic sex chromatin“ (PMSC) bezeichnet (Namekawa et al., 2006). Da in der Human Spermatogenese kein autosomales Retrogen die DDX3-Funktion während der Meiose übernehmen kann, muß in Human ein anderer Weg gefunden werden, um die Proteinexpression zu gewährleisten. Eine Möglichkeit besteht in der Synthese größerer mRNA-Mengen vor der Inaktivierung der Gonosomen und anschließender Speicherung in mRNP-Komplexen.

Für Vasa (= MVH „mouse VASA homologue“) wie belle (=DDX3), konnten sowohl in der Maus, als auch in *Drosophila*, wichtige Funktionen bei der post-transkriptionellen Regulation festgestellt werden (Castrillon et al., 2000, Noce et al., 2001; Johnstone et al., 2005). Die beiden Proteine sind in elektronendichten RNA-reichen, cytoplasmatischen Granula, Keimzell-Granula genannt, lokalisiert, die im Verlauf der Spermatogenese in verschiedenen Formen vorkommen (Parvinen, 2005; Kotaja & Sassone-Corsi, 2007; Onohara et al., 2010). Der diffusen Verteilung und der dadurch erschwerten Beschreibung verdanken solche Keimzell-spezifischen Zellkompartimente auch den Namen „nuage“ (französisch für Wolke). Dieser Begriff wird häufig als Überbegriff für alle Sorten von Keimplasma verwendet. In männlichen Keimzellen der Säugetiere wird der Begriff „nuage“ für prämeiotische elektronendichte Strukturen verwendet (Eddy, 1974; Onohara et al., 2010). Diese im Zytoplasma fibrösen und diffus verteilten RNP-Komplexe

werden in der Spermatogenese erstmals in pachytänen Spermatozyten sichtbar und kondensieren postmeiotisch in haploiden Keimzellen zu einem einheitlichen Zellkompartiment. Dieses postmeiotische Zellkompartiment wird, auf Grund der Anfärbbarkeit mit basischen Farbstoffen vergleichbar nukleärem Chromatin, als „Chromatoid body“ (CB) bezeichnet (Benda, 1891; Eddy, 1974; Onohara et al., 2010). Über zytoplasmatische Brücken, sogenannte Ringkanäle, kann sich der CB in einem aktiven Prozeß zwischen haploiden X- und Y-Spermatiden bewegen und so zu einem Materialaustausch, der genetisch unterschiedlichen, aber synchronisierten Spermatiden einer Spermatogenesewelle, beitragen (Ventelä et al., 2003; Dadoune et al., 2004; Parvinen, 2005). Dem CB werden Funktionen wie Translationskontrolle, mRNA-Speicherung, Remodellierung von RNP-Komplexen und die Unterdrückung von Transposons zugeschrieben (Arkov & Ramos, 2010). In somatischen Zellen sind analog zum CB „P-Bodies“ in der post-transkriptionellen Regulation von mRNAs beteiligt (Eulalio et al., 2007). Neuere Arbeiten zeigen, daß in CBs drei Proteinklassen, DEAD-Box Helikasen, Tudor-Domänen Proteine und MIWI/MILI-Proteine des RISC-Komplexes (RNA-induced silencing complex), wohl auch funktionell zusammenkommen (Kotaja & Sassone-Corsi, 2007; Arkov & Ramos, 2010). In *Drosophila* Keimzellplasma konnte eine Colokalisation von belle und vasa gezeigt werden (Johnstone et al., 2005). Die Bildung von stabilen DEAD-Box-Protein Homo- (Klostermeier & Rudolph, 2009) und Heterodimeren (Ogilvie et al., 2003) ist bekannt und könnte somit auch im CB zwischen VASA und DDX3X/Y möglich sein. Von DDX3X ist zudem bereits eine Beteiligung an somatischen mRNP-Komplexen bekannt (Merz et al., 2007).

1.5. Evolution der Säuger & Primaten

1.5.1. Evolution der Säugetiere

> »*Ich werde niemals gelten lassen, daß der Mensch, unter dem Vorwand, daß es eine Kluft zwischen ihm und den Tieren gibt, einen anderen Ursprung hätte*«[4]
> (Charles Darwin, 1838).

Die Stellung des Menschen im Stammbaum der Tiere hat die Menschheit seit jeher bewegt und wurde vielfach kontrovers diskutiert. Heute ist bekannt, daß die Säugetiere aus einer Gruppe amniotischer Landwirbeltiere namens Therapsida hervorgegangen sind, die als „säugetierähnliche Reptilien“ bezeichnet werden. Ihr Vorkommen war zwischen dem Unterperm[5] und der Unterkreide, d.h. vor ~ 280 bis 100 Millionen Jahren (Rubidge & Sidor, 2001).

---

[4] Vgl. de Bonis, L., (2001): S. 6.

[5] Einteilung der geologischen Erdzeitalter nach Walker & Geissman (2009).

Die erste große Aufspaltung innerhalb der Säugetiere erfolgt vor ~ 166 Mio. Jahren, als sich die Entwicklungslinien der eierlegenden Unterklasse Protheria, mit der einzig rezenten Ordnung Monotremata, und den lebendgebärenden Theria, bestehend aus Metatheria und Eutheria, trennten (Abb. 1-5; Bininda-Emonds et al., 2007). Die zweite Aufspaltung fand vor ~ 148 Millionen Jahren statt, als die Entwicklungslinien der Beutelsäuger (Metatheria) und der höheren Säugetiere (Eutheria) divergierten (Abb. 1-5; Bininda-Emonds et al., 2007).

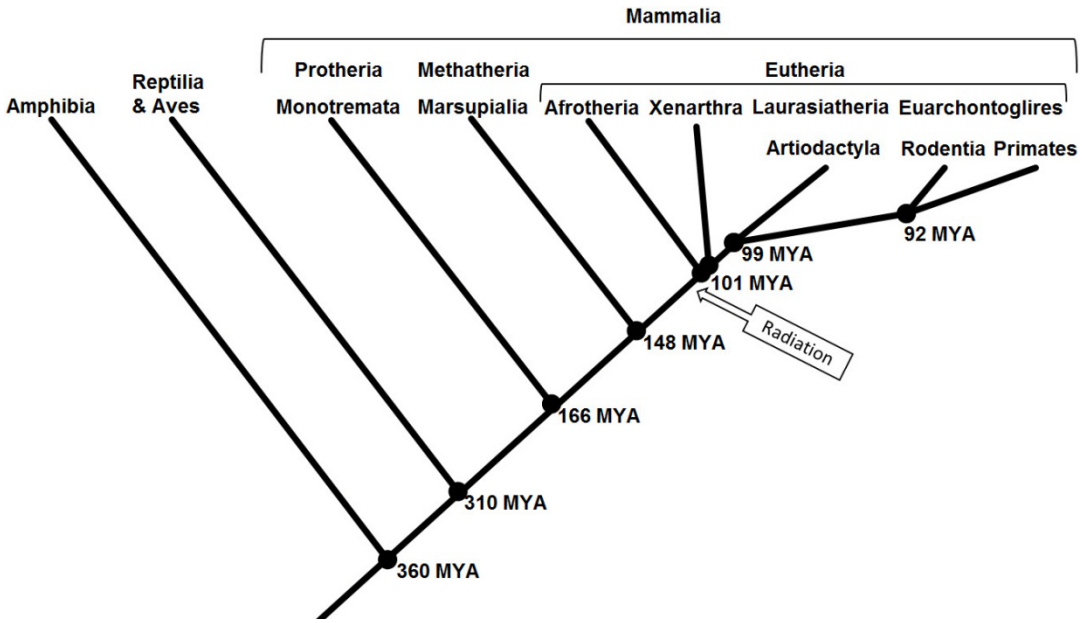


Abb. 1-5: Schematischer Wirbeltierstammbaum mit evolutionärer Zeittafel.
Zahlen geben die ungefähre Zeit seit Auftreten eines letzten gemeinsamen Vorfahren (LCA = last common ancestor) in Millionen Jahren (MYA = million years ago) an. Laurasiatheria ist ein systematisches Taxon, das die sieben Säugetierordnungen der Paarhufer (Artiodactyla), Raubtiere (Carnivora), Wale (Cetacea), Fledertiere (Chiroptera), Insektenfresser (Eulipotyphla), Unpaarhufer (Perissodactyla) und Schuppentiere (Pholidota) zusammenfaßt. Als Schwestertaxon gelten die Euarchontoglires, die aus den zwei Taxa Glires [Nagetiere (Rodentia) und Hasenartige (Lagomorpha)] und Euarchonta [Riesengleitern (Dermoptera), Primaten (Primates) und Spitzhörnchen (Scandentia)] zusammengesetzt sind. Aufspaltungszeitpunkte nach Hedges & Kumar (2003), Bininda-Emonds et al. (2007) und Janečka et al. (2007).

Die zeitliche Entstehungsgeschichte der Säuger und vor allem der Primaten kann durch die Datierung von Fossilien in Kombination mit einer Methode namens „Molekulare Uhr“ rekonstruiert werden. Die Idee einer molekularen Uhr, die Unterschiede zwischen Aminosäuresequenzen proportional zur Aufspaltungszeit der untersuchten Arten beschreibt, wurde erstmals von Zuckerhandl und Pauling (1965) vorgestellt. Anhand einer solchen molekularen Uhr und der Kenntnis der Mutationsraten der Proteine können Aufspaltungsereignisse von Arten auch ohne ausreichende Fossilien datiert werden, um so einen letzten gemeinsamen Vorfahren („last common ancestor“ = LCA) von zwei Entwicklungslinien vorherzusagen (v. Haesler & Liebers, 2003). Die Universalität der molekularen Uhr ist allerdings stark abhängig von der Evolutionsrate in proteinkodierenden Sequenzen, mit der neutrale Substitutionen auftreten (Kimura, 1983). In Proteinen, bei denen die meisten Aminosäuren für die Funktion notwendig sind, werden

schädliche, negative Substitutionen durch eine reinigende Selektion wieder aus der Population eliminiert. Die wenigsten Mutationen sind daher tatsächlich selektiv neutral. Die lineare molekulare Uhr wird über den Vergleich nicht-synonymer Substitutionen (K_A) zu synonymen Substitutionen (K_S) berechnet, wobei die Substitutionsrate an nicht-synonymen Positionen erwartungsgemäß geringer ist als an synonymen, da synonyme Substitutionen als beinahe selektionsneutral gelten und sich somit in einer Population etablieren können (Kimura, 1983; Li, 1993; Hedges & Kumar, 2003; v. Haesler & Liebers, 2003). Die molekulare Uhr wird aber auch durch intrinsische Faktoren wie Generationszeit (besonders bei Rodentia), Effizienz der DNA-Reparatur und Stoffwechsel, sowie extrinsische Faktoren wie Umwelteinflüsse, Anzahl der verwendeten Sequenzen, Art und Anzahl der Kalibrierungspunkte, beeinflußt, was zu unterschiedlichen Datierungen und Stammbäumen führen kann (v. Haesler & Liebers, 2003; Pearks Wilkerson et al., 2008). Zudem konnte auch gezeigt werden, daß X- und Y-homologe Gene nicht mit konstanten Raten evolvieren, wobei die Y-Kopien höhere Raten aufweisen (Ross et al., 2005; Pearks Wilkerson et al., 2008). Auch ist bekannt, daß unterschiedliche K_A- und K_S-Raten in unterschiedlichen Spezies existieren (Ross et al., 2005; Pearks Wilkerson et al., 2008). Dies gilt besonders für die Abstammungslinien der Rodentia und Primaten, für die eine erhöhte Evolutionsgeschwindigkeit proteinkodierender Gene festgestellt werden konnte (Mayr, 2005).

Die heutige Artenvielfalt der Eutheria entstand durch eine Radiation der Säugetierarten nach der Aufspaltung der Theria-Entwicklungslinie. Molekulare Daten datieren die Entstehung und Radiation der Eutheria zu Beginn der Oberkreide vor ~ 100 Millionen Jahren (Hedges & Kumar, 2003; Bininda-Emonds et al., 2007; Janečka et al., 2007). Ji et al. (2002) berichten von den frühesten Eutheriafossilien aus der Unterkreide vor ~ 125 Millionen Jahren. Die meisten Eutheria-Fossilfunde datieren allerdings erst nach dem Massenaussterben vor ~ 65,5 Millionen Jahre an der Kreide/Tertiär (K/T) Grenze, in dem auch die Dinosaurier ausstarben (Martin et al., 2007).

Die Eutheria werden in vier Überordnungen, Xenarthra (Nebengelenktiere), Afrotheria, Laurasiatheria und Euarchontoglires, eingeteilt, die vor ~ 100 Millionen Jahren, innerhalb von nur ~ 3 Millionen Jahren, entstanden (Hedges & Kumar, 2003; Bininda-Emonds et al., 2007). Die Primaten sind in der Überordnung Euarchontoglires, auch Supraprimates genannt, zu finden, die sich vor ~ 99 Millionen Jahren von den Laurasiatheria (z.B. das Rind) trennten (Abb. 1-5). Die nächste Aufspaltung auf dem Weg zu den Primaten erfolgte zwischen den Taxa Glires (Lagomorpha u. Rodentia) und Euarchonta (Dermoptera, Primates u. Scandentia) vor ~ 92 Millionen Jahren (Bininda-Emonds et al., 2007; Janečka et al., 2007). Die Aufspaltung der Euarchonta erfolgte wiederum in zwei Schritten. Erst erfolgte die Aufspaltung zwischen den

Scandentia (Spitzhörnchen) und Primatomorpha (Dermoptera u. Primates) und schließlich zwischen den Dermoptera (Riesengleitern) und den Primaten vor ~ 80 Millionen Jahren (Janečka et al., 2007). Andere molekulare Daten deuten darauf hin, daß die Entstehung der Primaten bis vor ~ 92 Millionen Jahren in die Oberkreide datieren könnte (Martin et al., 2007; Bininda-Emonds et al., 2007; Chatterjee et al., 2009; Wilkinson et al., 2011). Die Aufspaltungszeitpunkte sind allerdings nicht gesichert, da nur sehr wenige Fossilfunde die molekularen Daten unterstützen können und der älteste gesicherte Primaten Fossilfund auf ~ 55 Millionen Jahre datiert wurde (Martin et al., 2007). Als Ur-Primat stellt man sich ein Tier vor, das den heutigen Tupajas (Spitzhörnchen) ähnlich sehen könnte.

1.5.2. Evolution der Primaten

Während der Evolution der Primaten bis zur Gattung *Homo* erfolgten mindestens acht größere Aufspaltungen, die in Abbildung 1-6 schematisch dargestellt sind. Die Ordnung der Primaten werden zunächst in zwei Unterordnungen, Haplorrhini (Trockennasenaffen) und Strepsirrhini (Feuchtnasenaffen), eingeteilt. Alte Bezeichnungen für die Strepsirrhini sind Halbaffen oder Prosimiae und für die Haplorrhini „echte Affen“ oder Simiae. Diese zwei Entwicklungslinien trennten sich vor ~ 64 bis 80 Mio. Jahren (Nr. 1 in Abb. 1-6). Die phylogenetische Stellung der Koboldmakis (Tarsiiformes) ist nach wie vor umstritten. Sie werden entweder als Schwestertaxon der Anthropoidea innerhalb der Haplorrhini, oder als Schwestertaxon der Lemuren und Loris geführt. Chatterjee et al. (2009) ordnen sie als Schwestertaxon der Lemuren und Loris, mit einer Divergenzzeit von ~ 59 Mio. Jahren zu diesen, ein. Matsui et al. (2009) sehen sie als Schwestertaxon der Anthropoidea mit einer Divergenzzeit von 70 bis 77 Mio. Jahren von diesen. Die Aufspaltung der Lemuren und Loris erfolgte dann vor ~ 52 Mio. Jahren.

Die Aufspaltung der Catarrhini (Altweltaffen) und Platyrrhini (Neuweltaffen) innerhalb der Anthropoidea erfolgte vor ~ 40 bis 43 Mio. Jahren (Nr. 2, Abb. 1-6). Vor ~ 29 Mio. Jahren erfolgte die Trennung der Cercopithecoidea (geschwänzte Altweltaffen) und der Hominoidea (Zalmout et al., 2010). Die Trennung der großen Menschenaffen (Hominidea) und der kleinen Menschenaffen (Hylobatidae) erfolgte dann ~ vor 15 bis 17 Mio. Jahren. Die Familie der Hominiden besteht aus vier rezenten Gattungen mit den drei afrikanischen Gattungen (*Gorilla*, *Pan*, *Homo*) und der asiatischen Gattung (*Pongo*). Die Aufspaltung dieser geographisch getrennten Gattungen erfolgte vor ~ 11,3 bis 14 Mio. Jahren (Nr. 5 in Abb. 1-6). Die Gattung *Gorilla* divergierte vor ~ 6,4 bis 8,1 Mio. Jahren von der *Homo*/*Pan*-Entwicklungslinie. Die Abspaltung der Homininen, die die Gattung *Homo* und alle direkten Vorfahren seit Trennung der Gattung *Pan* umfassen, erfolgte vor ~ 5,4 bis 6 Mio. Jahren (Hedges & Kumar, 2003; Raaum et al., 2005; Yoder, 2007; Janečka et al., 2007; Chatterjee et al., 2009; Zalmout et al., 2010).

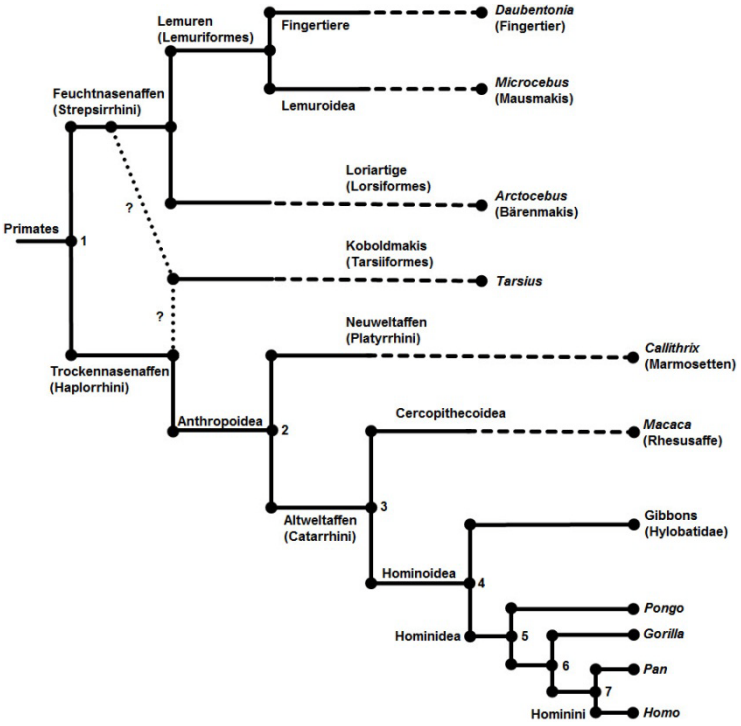


Abb. 1-6: Schematischer Stammbaum der Primaten.
Gezeigt werden nur rezente Taxa. Die Länge der Verbindungsstriche hat keine Aussage über die Divergenzzeit der verschiedenen Taxa. Die sieben Ziffern geben die größeren Aufspaltungsereignisse bis zur Gattung *Homo* wieder. **1** = Aufspaltung der Strepsirrhini - Haplorrhini vor ~ 64 - 80 MYA, **2** = Aufspaltung der Platyrrhini - Catarrhini vor ~ 40 - 43 MYA, **3** = Aufspaltung der Cercopithecoidea - Hominoidea vor ~ 29 MYA, **4** = Aufspaltung der Hylobatidae - Hominidae vor ~ 15 - 17 MYA, **5** = Aufspaltung afrikanische Hominidae - asiatischen Hominidae (*Pongo*) vor ~ 11,3 - 14 MYA, **6** = Aufspaltung der *Gorilla* - Pan/Homo vor ~ 6,4 - 8,1 MYA, **7** = Aufspaltung von *Pan* - Hominini vor ~ 5,4 - 6 MYA. Kladogramm erstellt nach Hedges & Kumar (2003), Raaum et al. (2005), Yoder (2007), Janečka et al. (2007), Chatterjee et al. (2009), Zalmout et al. (2010).

1.6. Evolution der Human Geschlechtschromosomen X und Y

In der Entwicklungslinie der Säugetiere ist der männliche Organismus das heterogametische Geschlecht mit zwei unterschiedlichen Gonosomen X und Y. Das weibliche Geschlecht ist homogametisch mit zwei X Chromosomen, die während der Meiose normal rekombinieren können. Die Human Gonosomen können während der Meiose, bis auf die Bereiche der pseudoautosomalen Regionen PAR1 (~ 2,6 Mb) und PAR2 (~ 320 kb) an den Chromosomenenden, nicht mehr rekombinieren und somit auch kein genetisches Material austauschen. Das männliche Y-Chromosom wird somit stets klonal vom Vater zum Sohn weitergegeben (Ellegren, 2011).

Das gängige Modell für die Entstehung der Human Gonosomen postuliert ein Proto-Gonosomenpaar autosomalen Ursprunges mit sukzessiv unterdrückter Rekombination der zwei divergierenden Gonosomen in der männlichen Meiose (Bull, 1983; Charlesworth et al., 1986; Charlesworth, 1996; Charlesworth et al., 2005). Lahn & Page (1999) nutzten diese Grundlage zur Berechnung gonosomaler Sequenzunterschiede in kodierenden X-Genen. Sie nahmen an, daß die Gonosomen X und Y, über die Jahrmillionen fehlender Rekombination, unabhängig voneinander Mutationen ansammeln würden, die die Gonosomen zunehmend verändern. Sie berechneten die K_A- zu K_S-Substitutionsraten und fanden in unterschiedlichen Loci vier

Divergenzklassen, was zu einem „Evolutionary Strata“ genannten Modell, mit Bereichen unterschiedlichen Entstehungsalters, führte (Abb. 1-7).

Die Entstehung der Human Geschlechtschromosomen setzte nach Trennung der Protheria–Theria Entwicklungslinien vor mehr als 150 Mio. Jahren ein (Veyrunes et al., 2008; Delbridge et al., 2009). Das älteste Stratum 1, auch als „X conserved region“ (XCR) bezeichnet, gilt als Rest des autosomalen Proto-X und bildet heute den langen Arm des X-Chromosoms Xq (Abb. 1-7).

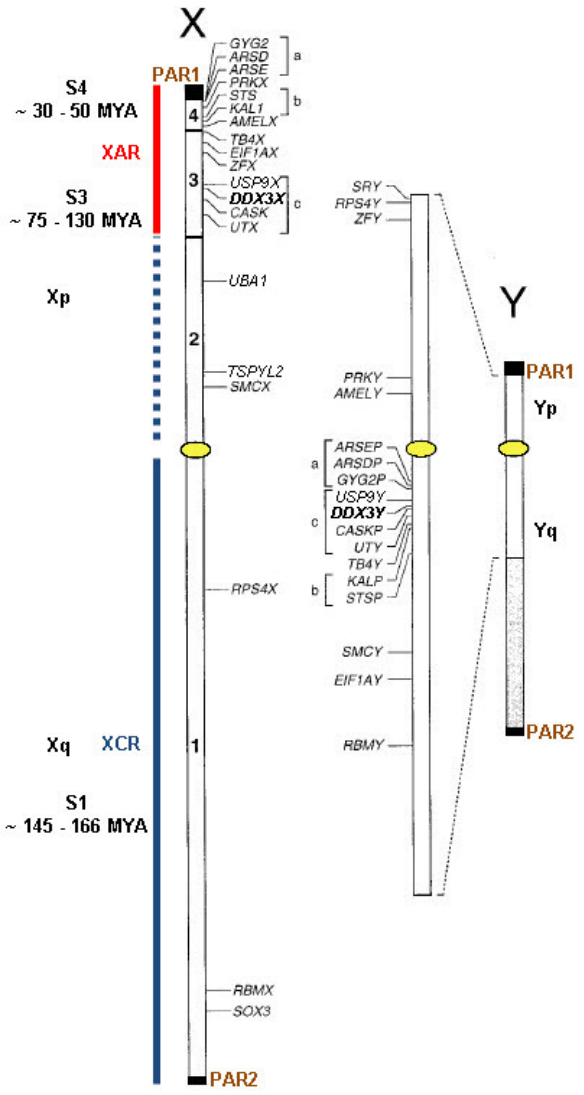


Abb. 1-7: Human Geschlechtschromosomen im schematischen Vergleich.
Zentromere sind als gelbe Ovale, PARs mit schwarzen Balken am Ende der Gonosomen dargestellt. Der Heterochromatin-Block in distal Yq ist schraffiert dargestellt. Die putative Grenze zwischen Stratum 1 und 2 soll zwischen den Genloci *SMCX* (= *KDM5C*) und *RPS4X* liegen und ist hier im Zentromer positioniert. Das Gen *UBA1* (= *UBE1X*) hat kein Human Y-Homolog. Eine potentielle Expansion von XCR in Xp ist durch die gestrichelte Linie angedeutet. Weitere Erläuterungen zur Graphik sind im Text beschrieben. *UTX* = *KDM6A*. Graphik verändert nach Lahn & Page (1999), Ross et al. (2005), Veyrunes et al (2008), Delbridge et al. (2009).

Der Zeitpunkt der Säuger-Gonosomenevolution wurde auf einen Zeitpunkt nach der Protheria–Theria Aufspaltung datiert, da für das Schnabeltier (*Ornithorhynchus anatinus*) Vogel-ähnliche Geschlechtschromosomen festgestellt werden konnten (Veyrunes et al., 2008). Die für Säugetiere typische männliche Geschlechts-differenzierung über das Gen *SRY*, das aus dem *SOX3* Gen in Stratum 1 entstanden ist, konnte für das Schnabeltier nicht identifiziert werden.

Nach Trennung der Metatheria, aber vor der Säugetierradiation, wurde dem X-Chromosom durch Translokation ein autosomaler Block, „X added region“ (XAR) genannt, zugefügt, der heute den kurzen Arm bildet (Delbridge et al., 2009). Zu diesem Zeitpunkt müssen XAR und YAR („Y added region“) noch rekombiniert haben (Wilson & Makova, 2009). Das Stratum 2 ist in seiner Entstehung noch umstritten und es ist nicht eindeutig, ob es sich hierbei um einen eigenen evolutionären Block handelt.

Für das von Lahn & Page (1999) in Stratum 2 kartierte Gen *UBE1X* (= *UBA1*) ist kein Human Y-Homolog vorhanden, so daß der K_S-Wert und der K_S/K_A-Quotient mit der Sequenz eines Neuweltaffen ermittelt wurden. Der ermittelte K_S-Wert für das Genpaar *SMCX/Y* in Stratum 2 ist deutlich unterschiedlich zu den drei anderen Strata 1, 3 und 4. Der K_S/K_A-Quotient fügt sich allerdings in das Spektrum der Stratum 1 und 3 Genpaare und stellt somit die Existenz dieses Stratum 2 in Frage.

Das Stratum 3 in der XAR ist ebenfalls noch vor der großen Radiation vor 75 bis 130 Mio. Jahren entstanden. Zu diesem Zeitpunkt bestand durch Y-chromosomale Rearrangements, wie Inversionen, bereits eine eingeschränkte Rekombinationsfähigkeit der Gonosomen. Das Gen *DDX3X* wurde anhand des K_S/K_A-Quotienten in Stratum 3 kartiert. *DDX3X* befindet sich in einem Genclusters mit [*USP9X–DDX3X–UTX*], welches in identischer Reihenfolge [*USP9Y–DDX3Y–UTY*] auf dem Human Y-Chromosom wiederzufinden ist. Die Syntenie, die Reihenfolge und Orientierung von Genen zueinander beschreibt, ist jedoch nicht komplett konserviert, da das Gen *UTY* durch eine Inversion anders orientiert ist. Die Integrität des AZFa-Locus konnte per FISH-Kartierung für alle Hominoiden bestätigt werden (Abb. 1-8; Wimmer et al., 2005). Die homologe Y-chromosomale Sxr^b-Region (sex-reversed) in Maus beinhaltet u.a. auch diese drei Gene in konservierter Reihenfolge (Mazeyrat et al., 1998; Affara, 2001). Da diese AZF-Gen-reihenfolge auch stabil in Schwein (Quilter et al., 2002) und Katze (Murphy et al., 1999; Pearks Wilkerson et al., 2008) identifiziert werden konnte, folgern Wimmer et al. (2005), daß die konservierte AZFa-Region, im Gegensatz zur allgemeinen Dynamik der Säuger Y-Chromosomen, in den meisten oder allen Eutheria eine kritische Funktion für die Fertilität besitzt.

Das jüngste Stratum 4, welches in Nähe zur PAR1 lokalisiert ist, entstand vor ~ 30 bis 50 Mio. Jahren innerhalb der Haplorrhini, nach der Aufspaltung der Primaten in die Entwicklungslinien der Haplorrhini und Strepsirrhini (siehe Gabelung Nr. 1 in Abb. 1-6). Kohn et al. (2004) schlagen anhand von vergleichenden Human–Huhn Genkartierungen ein modifiziertes Strata-Modell vor. Die Lahn & Page (1999) Strata 1 und 2 werden zu dem neuen Stratum 1 (Xp11.2–Xqter = XCR) und die alten Strata 3 und 4 werden zu dem neuen Stratum 3 (Xp11.2–Xpter = XAR). Hinzu kommen die neu definierten genreichen Strata 2a (Xp11.2) und 2b (Xq28). Die Arbeiten von Veyrunes et al. (2008) und Delbridge et al. (2009) basieren auf diesem modifizierten Strata-Modell von Kohn et al. (2004). Festzuhalten ist, daß die divergierende Entwicklung der Geschlechtschromosomen vor ~ 130 Mio. Jahren mit Stratum 3 begannen und zu den heute sichtbaren Unterschieden in Größe (X-Chr. = 155,3 Mb; Y-Chr. = 59,4 Mb; „Human Genome Assembly“ GRCh37) und Genausstattung der Human Geschlechtschromosomen führte.

Die Evolution innerhalb des Y-Chromosoms der Primaten ist sehr dynamisch, wie der Vergleich in fünf Hominidenspezies zeigt (Abb. 1-8). Seit Trennung der Gattung *Pongo* vor 11,3 bis 14 Mio. Jahren sind in den unterschiedlichen Spezies mehrere speziesspezifische Duplikationen und Inversionen aufgetreten (Gläser et al., 1998; Wimmer et al., 2005). Selbst in den am nächsten verwandten Gattungen *Homo* und *Pan* sind große Unterschiede in Genausstattung (Tab. 1-1) und Chromosomenstruktur feststellbar, die durch speziesspezifische Transpositionen, Inversionen und Deletionen entstanden sind (Abb. 1-9; Kehrer-Sawatzki & Cooper, 2007a, 2007b; Hughes et al., 2010).

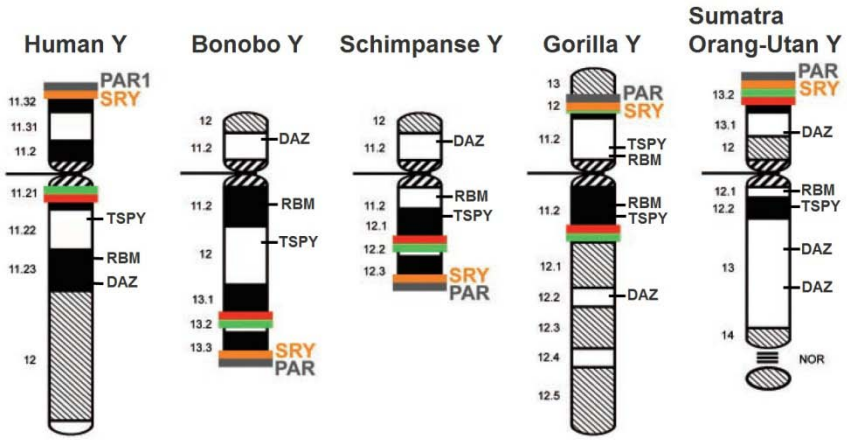


Abb. 1-8: Y-Chromosomen fünf verschiedener Hominiden in Bezug zur AZFa Lokalisation.
Die AZFa-Lokalisationen wurde durch FISH-Kartierung ("fluorescence *in situ* hybridisation") Humaner PAC-Klone erstellt. Die Human AZFa-Region wird durch den proximalen grünen Balken und den distalen roten Balken im Y-Idiogramm in proximal Yq markiert. Human PAR1 ist als grauer Balken in Yp dargestellt. Die Non-Human Primaten verfügen über nur eine PAR, die ortholog zu PAR1 ist. Für Schimpanse, Bonobo und Orang-Utan konnte eine AZFa-Lokalisation nahe der jeweiligen PAR festgestellt werden. Nur in Gorilla konnte eine Lokalisation vergleichbar zu Human auf dem anderen Y-Chromosomenarm kartiert werden. Die AZFa-Orientierung ist in Schimpanse, Bonobo und Gorilla umgekehrt. Zusätzlich zur AZFa- und SRY-Lokalisation sind die Genloci für *TSPY*, *RBM* und *DAZ* eingezeichnet. Für diese Gene konnten ebenfalls unterschiedliche Positionen, relative Orientierungen und Anzahlen identifiziert werden. NOR = „Nucleolus organiser region". Nach Wimmer et al. (2005) und Gläser et al. (1998).

Auf molekularer Ebene läßt sich das heutige Human Y-Chromosom in sechs Sequenzklassen einteilen (Abb. 1-9; Skaletsky et al., 2003). Es beinhaltet 78 proteinkodierende Gene, die für 27 unterschiedliche Proteine kodieren (Tab. 1-1). Bis auf die terminalen PARs sind die Sequenzen in ihrer Zusammensetzung und Reihenfolge Y-spezifisch und werden als „male-specific region of the Y-chromosome" (MSY) bezeichnet. Die zuvor gebräuchliche Bezeichnung „non-recombining portion of the Y" (NRY) wird nicht mehr verwendet, da Y-Y Rekombinationsereignisse regelmäßig auftreten (Skaletsky et al., 2003). Der Begriff MSY ist allerdings auch mißverständlich, da mit der Bezeichnung suggeriert wird alle Sequenzen seien Y-spezifisch und hätten nicht den gemeinsamen Ursprung als Proto-X und Proto-Y. MSY umfaßt 95% des 59,4 Mb großen Human Y-Chromosoms.
Die ältesten Y-Sequenzbereiche, die die Überreste der Human Proto-Gonosomen darstellen, werden „X-degenerate" genannt.

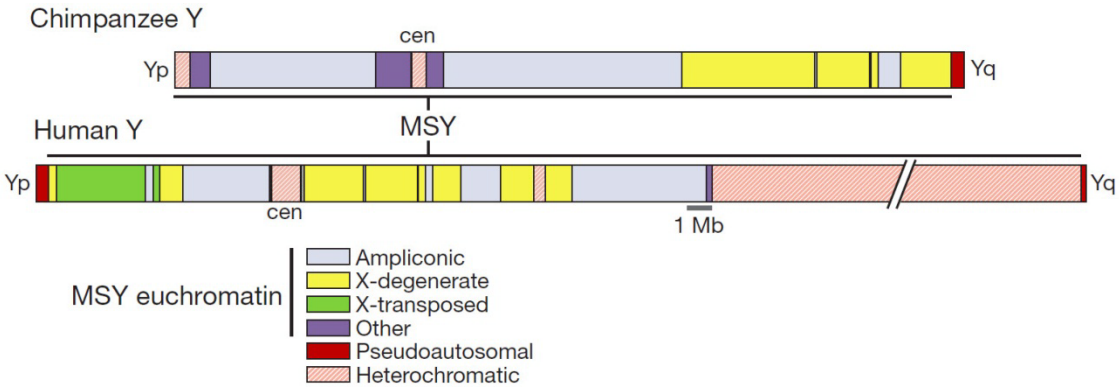


Abb. 1-9: Schematischer Vergleich des Human und Schimpansen Y-Chromosoms.
Gezeigt sind sechs unterschiedliche Sequenzklassen. Vier Klassen etablieren den euchromatischen Teil (~ 23 Mb) des Human Y-Chromosoms. Das Heterochromatin umfaßt einen ~ 40 Mb langen Block in Yq und das Zentromer mit ~ 1 Mb Länge. Die „X-transposed" Sequenzklasse (grün markiert) ist humanspezifisch und hat eine kombinierte Länge von 3,4 Mb. Die Amplikone umfassen kombiniert 10,2 Mb, was ~ 25% der MSY ausmacht. Die Klasse „other" bezeichnet single-copy Sequenzen, die weder „X-degenerate" (gelb markiert) oder „X-transposed" sind. Die Human „Pseudoautosomalen Regionen" PAR1 und -2 an den äußersten Enden sind in rot eingezeichnet. Der Schimpanse, sowie die meisten anderen Säugetiere, verfügen nur über eine PAR (Bondy & Cheng, 2009). Modifiziert nach Hughes et al. (2010).

In diesen Blöcken kommen ausschließlich singel-copy Gene und Pseudogene vor. Es konnten 16 proteinkodierende Gene in „X-degenerate" festgestellt werden, die 27 Genen auf dem Human X-Chromosom entsprechen. In den Schimpansen „X-degenerate" Sequenzen sind 12 der 16 Gene erhalten. Vier dieser Gene, u.a. *USP9Y*, haben negative Mutationen, die nicht mehr zu einem Protein führen (Tab. 1-1). Die zwei Blöcke, die als „X-transposed" bezeichnet werden, sind humanspezifisch und nicht im Schimpansen Y-Chromosom zu detektieren (Hughes et al., 2010). Sie sind das Ergebnis eines einzelnen X zu Y Transpositionsereignis, mit nachfolgender Inversion in der Human MSY, die diesen Xq21 homologen Block in zwei Teile spaltete (Skaletsky et al., 2003). Es sind zwei proteinkodierende Gene in „X-transposed" lokalisiert (Tab. 1-1).

Die dritte euchromatische Klasse wird als „Amplikon" bezeichnet und ist sowohl auf dem langen als auch auf dem kurzen Y-Arm lokalisiert. Die Amplikone sind in sieben Blöcken repetitiver und sehr homologer DNA organisiert, die meist als Palindrome angeordnet sind (Skaletsky et al., 2003). Das Schimpansen Y-Chromosom verfügt ebenso über mehrere Amplikonblöcke, die allerdings fast doppelt so groß sind wie in der Human MSY und in 19 Palindromen (Human acht) angeordnet sind (Hughes et al., 2010). Die Amplikone beinhalten vor allem multi-copy Genfamilien, die exklusiv oder prädominant in Hodengewebe exprimiert werden. In den Human MSY Amplikonen sind ~ 60 proteinkodierende Gene lokalisiert, die lediglich neun Genfamilien zugeordnet werden (Skaletsky et al., 2003). Für drei Genfamilien (*TSPY*, *VCY*, *RBMY*) konnte jeweils ein single-copy X-Homolog identifiziert werden (Skaletsky et al., 2003; Lau et al., 2007).

Zwei Genfamilien (*CDY*, *DAZ*) haben auto-somale single-copy Homologe (Skaletsky et al., 2003).

Tab. 1-1: MSY Gene und Genfamilien in Human und Schimpansen Y-Chromosom.
Aufgelistet sind funktionelle MSY Gene. Der AZFa-Locus ist grau unterlegt. Verändert nach Skaletsky et al. (2003), Lau et al. (2007) und Hughes et al. (2010).

MSY Sequenzklasse	Gen *Homo* Y	Anzahl intakte Kopien *Homo* Y	X-Homolog *Homo*	autosomales Homolog *Homo*	Anzahl intakte Kopien *Pan* Y
X-degenerate	*SRY*	1	*SOX3*	-	1
	RPS4Y1	1	*RPS4X*	-	1
	ZFY	1	*ZFX*	-	1
	AMELY	1	*AMELX*	-	1
	TBL1Y	1	*TBL1X*	-	-
	PRKY	1	*PRKX*	-	1
	USP9Y (= *DFFRY*)	1	*USP9X* (= *DFFRX*)	-	-
	DDX3Y	1	*DDX3X*	-	1
	UTY	1	*UTX* (= *KDM6A*)	-	1
	TMSB4Y	1	*TMSBX*	-	-
	NLGN4Y	1	*NLGN4X*	-	1
	CYorf15A	1	*CXorf15* (= *TXLNG*)	-	1
	CYorf15B	1	*CXorf15* (= *TXLNG*)	-	-
	SMCY (= *KDM5D*)	1	*SMCX* (= *KDM5C*)	-	1
	EIF1AY	1	*EIF1AX*	-	1
	RPS4Y2	1	*RPS4X*	-	1
Σ		**16**			**12**
Amplikon	*TSPY*	~ 35	*TSPYL2* (= *TSPX*)	-	6
	VCY	2	*VCX*	-	2
	XKRY	2	-	-	-
	CDY	4		*CDYL*	5
	HSFY	2	-	-	-
	RBMY	6	*RBMX*	-	6
	PRY	2	-	-	-
	BPY2	3	-	-	2
	DAZ	4	-	*DAZL*	4
Σ		**~ 60**			**25**
X-transposed	*TGIF2LY*	1	*TGIF2LX*	-	-
	PCDH11Y	1	*PCDH11X*	-	-
Σ		**2**			**0**
Σ		**~ 78**			**37**

Insgesamt sind von 20 MSY Genen oder Genfamilien X-Homologe bekannt (Skaletsky et al., 2003; Vogt, 2006). Die NAHR Y-Y Rekombinationsereignisse finden zwischen den homologen und repetitiven Sequenzen statt und haben zu den sichtbaren Unterschieden der Human und Schimpansen Y-Chromosomen geführt. Ein möglicher Faktor der zu den spezifischen Y-Chromosomen führte ist das unter-schiedliche Reproduktionsverhalten von Mensch und Schimpanse.

In der heutigen *Homo sapiens* Population ist jedoch nicht nur „Das Human Y-Chromosom“ vorhanden, sondern es existieren regional unterschiedlich verteilte Y-Chromosomen[6],

[6] Vgl. „Y-DNA Haplogroup Tree 2011“ der ISOGG (International Society of Genetic Genealogy).

Haplogruppen genannt, die durch Variationen in den MSY-Sequenzen definiert werden (Jobling & Tyler-Smith, 2003). Es werden 20 Haupt-Haplogruppen (A bis T), mit weiteren Subgruppen, unterschieden (Chiaroni et al., 2009). Die Y-Chromosom Referenzsequenz (GenBank Acc.No. NC_000024), mit den in Tab. 1-1 aufgeführten Genen, entspricht der Haplogruppe R (Fernandes et al., 2004). Die unterschiedlichen Haplogruppen können Einfluß auf die Fertilität der Träger haben. Der patrilineare LCA wird als „Y-chromosomaler Adam" bezeichnet und wird in der Evolution des anatomisch modernen Menschen vor ~ 100 - 200 Mio. Jahren in Afrika vermutet (Gibbons, 1997; Tattersall, 2009; Blum & Jakobsson, 2011).

1.7. Transkriptionskontrolle in Eukaryoten

1.7.1. Der Core-Promoter

Für die eukaryotische Gentranskription durch die RNA-Polymerase II (RNAPII) werden bestimmte Sequenzen auf der DNA (*cis*-regulative Elemente) benötigt, an die *trans*-regulative Elemente, wie die Polymerase und Genregulatorproteine (allgemeine u. spezifische Transkriptionsfaktoren) binden können. Die Core-Promoter genannte Region beinhaltet die Transkriptionsstartstelle (TSS) eines Genes und nimmt somit eine zentrale Rolle bei der Genaktivierung ein. Der Core-Promoter ist als die minimale DNA-Region definiert, die in der Lage ist, eine basale, Aktivator-unabhängige Transkription durch die RNAPII *in vitro* zu dirigieren (Gross & Oelgeschläger, 2006). Typischerweise definiert man die Region jeweils 40 Basenpaare (bp) stromaufwärts und abwärts der TSS als Core-Promoter.
Die Core-Promoter Region beinhaltet meist kurze DNA Sequenzen von 3 bis ~ 10 Nukleotiden, den Core-Promotermotiven, an die Faktoren des Präinitiationskomplexes (PIC) zur Initiation der RNA-Synthese binden können. Der PIC setzt sich aus den „basal" oder „general transcription factors" (GTF) genannten Untereinheiten TFIIA, TFIIB, TFIID, TFIIE, TFIIF und TFIIH, sowie der DNA-abhängigen RNAPII zusammen, die für die Transkription proteinkodierender Gene in Eukaryoten zuständig ist. Einmal etabliert, kann der PIC mehrfach nacheinander die Transkription des betreffenden Genes initiieren (Morachis et al., 2010).

Die entscheidende Komponente für die DNA-Erkennung und DNA-Bindung des PIC ist dabei der TFIID-Multiproteinkomplex (Abb. 1-10), der aus TBP (TATA-Box binding Protein) und bis zu 14 weiteren Untereinheiten, den TAFs (TBP-associated factor), zusammengesetzt sein kann (Thomas & Chiang, 2006). Es wird angenommen, daß TFIID eine Nukleosomen-ähnliche Struktur bei der Interaktion mit der DNA annehmen kann, da in neun von 14 TAFs die „Histone-

fold Domain“ (HFD) zu finden ist, wie sie auch in den Histonen H2A, H2B, H3 und H4 vorkommt (Shao et al., 2005).

Neben TBP nehmen auch die TFIID-Untereinheiten TAF1, TAF2, TAF4b, TAF6, TAF9 und TAF12 spezifisch Kontakte zur DNA auf (Thomas & Chiang, 2006; Sengupta et al., 2009).
Die Bindungspolarität des PIC an den Core-Promoter wird von RNAPII zusammen mit TBP und den TFIIA- und TFIIB-Komplexen gewährleistet (Kays & Schepartz, 2000). Für die Bindung an die DNA ist die RNAPII zusammen mit TFIIF zuständig. Für das Entwinden der DNA kooperiert die RNAPII mit der Helikase TFIIH (Seila et al., 2008; Thomas & Chiang, 2006; Kays & Schepartz, 2000).

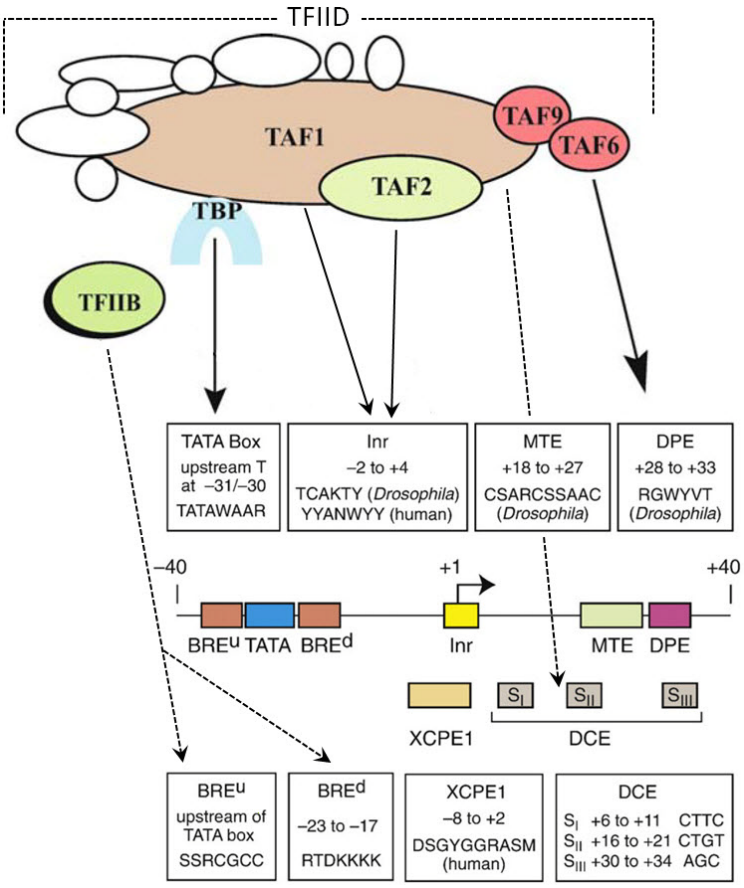


Abb. 1-10: Schematische Darstellung verschiedener Core-Promotermotive und deren Erkennung durch Untereinheiten des Transkriptionskomplexes.
Für TFIID werden nur die Untereinheiten (TAFs) benannt, die an der Erkennung und Bindung der Core-Elemente beteiligt sind. Die RNAPII ist nicht dargestellt. BRE = „TFIIB recognition element“, TATA = TATA-Box, Inr = Initiator, MTE = „Motif Ten Element“, DPE = „Downstream Promoter Element“, DCE = „Downstream Core Element“, TFII = „General Transcription Factor“ für die RNA-Polymerase II. Eine Liste der Core-Promotermotive und deren Erkennung durch Untereinheiten des PIC-Komplexes sind in Tab. 1-2 angegeben. Nach Thomas & Chiang (2006) und Juven-Gershon & Kadonaga verändert (2010).

Sandelin et al. (2007) und Juven-Gershon et al. (2008) unterscheiden zwei Core-Promoter Klassen, den fokussierten und den breiten Typus. Der fokussierte Core-Promoter ist meist auf eine bestimmte TSS konzentriert und ist häufig mit einer funktionellen TATA-Box und gewebespezifischer Expression korreliert (Tora & Timmers, 2010). Diese Form ist allerdings eher die Ausnahme unter den Vertebraten Core-Promotern, stellt aber offensichtlich die evolutionär ältere Variante dar, die von den Archaeen bis zum Menschen konserviert ist (Juven-Gershon et al., 2008).

Der breite Typus kann multiple TSS über 50 bis zu 100 Nukleotide verteilt aufweisen und ist bei Vertebraten häufig in Promoterregionen mit assoziiertem CpG-Island anzutreffen. Bis zu 50% der Human Promoterregionen sind mit CpG-Islands assoziiert, die häufig mit konstitutiv exprimierten, sogenannten „Housekeeping Genen", korreliert sind (Sandelin et al, 2007). Für beide Core-Promotertypen wird ein grundlegend unterschiedlicher Transkriptionsinitiations-mechanismus vermutet, da der Bindungsmodus des TFIID-Komplexes an einen TATA–Inr Core-Promoter andere DNA-Protein-Bindungen und DNA-Biegungen hervorruft, als die Bindung an einen TATA-losen Core-Promoter mit CpG-Island (Juven-Gershon & Kadonaga, 2010; Tora & Timmers, 2010). Mischformen zwischen beiden Typen, in denen in einem breiten Core-Promoter eine bestimmte TSS präferiert wird, sind ebenfalls bekannt (Sandelin et al., 2007). Mehrere TSS können auch in einem fokussierten Core-Promoter vorkommen, wenn mehrere starke Core-Elemente, wie z.B. TATA-Box und Initiatormotiv (Inr), um die PIC-Positionierung konkurrieren (Smale & Kadonaga, 2003; Ponjavic et al., 2006).
Die wichtigsten Core-Promotermotive für die Positionierung des PIC sind die TATA-Box und das Initiatormotiv (Tab. 1-2).

Tab. 1-2: Liste bekannter Core-Promotermotive.
Angabe der relativen Position zur TSS, der jeweiligen Konsensussequenz und der PIC Untereinheit, die das Element bindet. Die TSS (+1) Position ist in den entsprechenden Motiven unterstrichen. Das XCPE1 kommt in ~ 1% der humanen Promoter, hauptsächlich in CpG-Islands, vor. Nach Kutach & Kadonaga (2000), Jin et al. (2006), Juven-Gershon et al. (2008), Anish et al. (2009) und Theisen et al. (2010).

Core-Promoter Element	Position in Bezug auf TSS (+1)	Konsensussequenz 5´ – 3´	Bindung durch PIC-Untereinheit
BRE^u (TFIIB recognition element, upstream)	Direkt vor TATA	SSRCGCC (Säuger)	TFIIB
TATA-Box	-31 bis -24	TATAWAAR (Säuger)	TBP
BRE^d (TFIIB recognition element, downstream)	Direkt hinter TATA	RTDKKKK	TFIIB
Inr (Initiator)	-2 bis +5	YYANWYY (Human)	TAF1-TAF2-TBP Subkomplex
MTE (Motif Ten Element)	+18 bis +29	CSARCSSAAC (Drosophila) SVAGCSSRGCGS (Säuger)	TFIID
DPE (Downstream Core Promoter Element)	+28 bis +34	RGWYVT (Drosophila) DSWYVY (Drosophila)	TAF6 & TAF9
DCE (Downstream Core Element)	S1: +6 bis +11 S2: +16 bis +21 S3: +30 bis +34	S1: CTTC S2: CTGT S3: AGC	TAF1
XCPE1 (X Core Promoter Element 1)	-8 bis +2	DSGYGGRASM (Human)	TFIID oder freies TBP + TFIIB + Mediator-Komplex
XCPE2 (X Core Promoter Element 2)	-9 bis +2	VCYCRTTRCMY	TFIID oder freies TBP + TFIIB + Mediator-Komplex
GCG-Motiv (eventuell Säuger-Pendant zu MTE und DPE)	+20 / +30	GCG	unbekannt
Bridge-Element (Bereich 1 überlappt mit MTE, Bereich 2 mit DPE)	1) +18 - +22 2) +30 - +33	1) SVAGC 2) WYVT	TFIID

Es sind die einzigen Elemente, die ohne Hilfe weiterer Core-Promotermotive, wie z.B. DPE, BRE und MTE und ohne Unterstützung assoziierter Aktivator-TFs in der Lage sind, den PIC korrekt auf die DNA und an die TSS zu rekrutieren.

Die TATA-Box kommt in etwa 20% der Human Core-Promoterregionen als entscheidendes Core-Promotermotiv vor. Sie befindet sich in der Regel, im Abstand von -28 bis -34 bp (erstes T von TATA), mit Präferenz zu -30/-31 bp, stromaufwärts der TSS und wird von dem Faktor TBP erkannt (Smale & Kadonaga, 2003). Eine Besonderheit der TBP-Bindung an die DNA stellt die Interaktion mit der kleinen Furche der DNA dar, was, wie z.B. bei den TFs der SOX-Familie, zu einem Biegen der DNA führt. Die maximale DNA-Biegung um 90° kann allerdings nur bei kanonischer und funktioneller TATA-Box zu einer räumlichen Veränderung der Core-Promoterregion führen (Tora & Timmers, 2010). Die Funktionalität der TATA-Box wird also durch die Bindungsaffinität zu TBP bestimmt und stellt somit eine Qualitätskontrolle der Transkriptionsinitiation dar. Bindungsstudien von TBP an verschiedene TATA-Sequenzen haben ergeben, daß C:G oder G:C Basenpaare an allen Positionen außer 2, 4 und 5 toleriert werden können (Smale & Kadonaga, 2003). Aus den strukturellen Studien resultierten folgende Bindungsstärken für jedes einzelne Nukleotid der Konsensussequenz „TATAWAAR“[7]:
$[T >> c > a{\sim}g/A >> t/T >> a{\sim}c/A >> t/T >> a/A >> g > c{\sim}t/A{\sim}T > g > c/G{\sim}A > c{\sim}t]$.
Die Autoren betonen, daß diese Definition lediglich vorhersagt, wie gut jedes Nukleotid an jeder Position toleriert werden kann und nicht, ob TBP tatsächlich an eine solche Sequenz bindet. Smale & Kadonaga (2003) gehen nicht näher darauf ein, wie viele „schwächere“ Nukleotide pro TATA-Motiv toleriert werden können. Tokusumi et al. (2007) sprechen von einem Mismatch pro TATA-Motiv, beziehen sich aber auf die Konsensussequenz „TATAWAAR“. Delgadillo et al. (2009) berichten, daß Substitutionen in der Konsensussequenz zu einer verminderten Biegefähigkeit der DNA durch TBP führen, was wiederum zu einer erhöhten Verweildauer des TBP/TFIID-Komplexes auf der DNA führt. Es scheint allerdings möglich, daß TBP funktionell aktiv bleiben kann, ohne die ideale, stabile Sattelform, bei der Bindung an nicht-kanonische TATA-Boxen, anzunehmen (Smale & Kadonaga, 2003; Tora & Timmers, 2010).

Nicht-kanonische TATA-Boxen, die in ~ 70% der TATA-Box regulierten Gene vorkommen, haben häufig zur Folge, daß zusätzliche Core-Promotermotive zur Stabilisierung des PIC-Transkriptionskomplexes benötigt werden. Es ist bekannt, daß identische Mutationen in TATA-Boxen unterschiedlicher Gene zu sehr verschiedenen Effekten der Promoteraktivität führen können (Savinkova et al., 2009). Darüber hinaus haben TATA-Box flankierende Sequenzen

[7] Die verwendeten Nukleotidbezeichnungen entsprechen dem „Ambiguity Code“ der IUPAC und IUBMB („International Union of Pure and Applied Chemistry“ und „International Union of Biochemistry and Molecular Biology”).

eine entscheidende Funktion bei der DNA/TBP-Interaktion, so daß von einer funktionell spezifischen Umgebung für jede TATA-Box ausgegangen wird. Im Unterschied zu der TATA-Box weisen die flankierenden Regionen häufig CG-reichere Sequenzen mit höherem Schmelzpunkt, als die TATA-Box selbst, auf. Dieser Unterschied ermöglicht es dem TBP die TATA-Box effizient zu erkennen (Savinkova et al., 2009).

Das Initiator-Motiv umgibt direkt die TSS, die in Position drei im Initiator lokalisiert ist (Smale & Baltimore, 1989; Yang et al., 2007). Das dritte Nukleotid, im fokussierten Core-Promoter meist ein Adenin, wird daher als +1 Position bezeichnet – d.h. es ist das erste Nukleotid der mRNA. Das Inr wird von TAF1 und TAF2 erkannt. Häufig befindet sich im breiten Core-Promotertyp anstelle eines vollständigen Initiatormotives nur ein konserviertes Y-R-Dinukleotid[8] an der -1/+1 Position. Die Forderung eines Purins an Position +1 ist sehr stringent, das Vorhandensein eines Adenins ist dagegen nicht universell (Sandelin et al., 2007).
In TATA-gesteuerten Core-Promoter befindet sich die TATA-Box präferentiell an Position -30/-31 zu dem minimalen Y-R Motiv. Bei größerer Distanz (-32 bis -34) sind an den Positionen -2 bis -4 bevorzugt Pyrimidine anzutreffen, um die Entstehung neuer TSSs in idealerer Entfernung zu vermeiden (Ponjavic et al., 2006). Laut einer *in-silico* Studie von 2005, ist das Inr-Motiv, mit einer Anwesenheit von ~ 50% in den untersuchten Promotersequenzen, das am häufigsten auftretende Core-Promotermotiv in Human Promoterregionen (Gershenzon & Ioshikhes, 2005).

Das „TFIIB Recognition Element“ (BRE), welches durch den GTF „TFIIB“ erkannt wird, wurde ursprünglich in einer Untergruppe von TATA-Box-Promotern identifiziert (Lagrange et al., 1998; Juven-Gershon et al. 2008). Die funktionelle TFIIB-Bindung an BRE^u und BRE^d scheint nur in Abhängigkeit der TBP-Bindung an die TATA-Box zu erfolgen (Deng & Roberts, 2007). Je nach Promoterkontext kann die Wirkung sowohl synergistisch, als auch antagonistisch auf die Funktion der TATA-Box ausfallen (Lagrange et al., 1998; Deng & Roberts, 2005, 2007).

Die Konsensussequenzen von DPE-Motiv (Downstream Core Promoter Element) und MTE (Motif Ten Element) wurden eingehend in *Drosophila* untersucht. Funktionell vergleichbare Core-Promotermotive konnten nachfolgend ebenfalls in Säuger Core-Promoterregionen identifiziert werden (Juven-Gershon et al., 2008).
DPE-Motiv und MTE sind nur in Kombination mit dem Initiatormotiv funktionell, da TFIID kooperativ an Inr und DPE/MTE bindet (Gross & Oelgeschläger, 2006; Juven-Gershon et al., 2006a; Juven-Gershon & Kadonaga, 2010). Das DPE wird von TAF6 und TAF9 erkannt, und kann, wie auch das MTE, in TATA-losen Promoter das Inr-Motiv bei der Rekrutierung des PIC unterstützen (Goodrich & Tjian, 2010). DPE-abhängige Core-Promoter (Inr + DPE) sind,

[8] Y-R = Pyrimidin-Purin Dinukleotid (CA, TA, CG, TG)

gegenüber TATA-abhängigen, unterschiedlich empfänglich für Signale von Aktivator- oder Repressor-TFs (Kadonaga, 2002). So können Aktivator-TFs für eine DPE-abhängigen Core-Promoter als Repressor auf einen TATA-abhängigen, und umgekehrt, wirken (Kadonaga, 2002).

Die stringente hexamere DPE-Konsensussequenz „RGWYVT“ wurde über Bindungsstudien in *Drosophila* bestimmt (Burke & Kadonaga, 1997; Sengupta et al., 2009). Kutach und Kadonaga (2000) konnten zeigen, daß das DPE-Motiv „RGWYVT“ nur für eine Untergruppe von DPE-Motiven zutreffend ist und damit nicht alle funktionellen DPEs erfasst werden. Eine allgemeingültigere DPE-Konsensussequenz wurde als „DSWYVY“ definiert.

Für das MTE, das direkt stromaufwärts des DPE-Motives positioniert ist, konnte die bindende TFIID-Untereinheit bisher noch nicht definiert werden. Die „original“ MTE-Konsensussequenz „CSARCSSAACGS“ wurde in *Drosophila* bestimmt (Lim et al., 2004). Die für Säuger Core-Promoterregionen angepaßte MTE-Konsensussequenz „SVAGCSSRGCGS“ weicht geringfügig von der Originalsequenz ab (Jin et al., 2006).

Obwohl MTE und DPE unabhängig voneinander in Inr-abhängiger Weise funktionieren, gibt es eine Verbindung beider Elemente, die als zweiteiliges „Bridge Element“ definiert wurde (Theisen et al., 2010). Das „Bridge Element“ überlappt mit MTE und DPE, bzw. besteht aus den kritischen Bereichen, +18 bis +22 („SVAGC“, MTE) und +30 bis +33 („WYVT“, DPE), dieser zwei Elemente.

Bekannte Core-Promoterelemente zur Transkriptionsinitiation sind nicht für jedes Gen zu finden. Somit muß auch von alternativen Core-Promotermodellen, mit anderen PIC-Rekrutierungs- und Transkriptionsinitiationsmöglichkeiten ausgegangen werden. In breiten Core-Promoterregionen sind oftmals keine kanonischen Core-Promoterelemente vorhanden und an der TSS-Position ist häufig ein G-G-Dinukleotid zu beobachten (Sandelin et al., 2007). Hier stellt sich die Frage, wie der TFIID-Komplex rekrutiert und stabilisiert wird? Eine mögliche Antwort ist die direkte Proteininteraktion von TAFs mit aktivierenden TFs, wie SP1 oder CREB (Gill et al., 1994; Asahara et al., 2001; Ohler & Wassarman, 2010). Eine weitere Möglichkeit zur PIC-Positionierung ohne starke Core-Promoterelemente stellt die Interaktion von TAFs mit bestimmten Chromatinmodifikationen, wie trimethylierten (H3K4me3) und acetylierten Histonen (H3K9ac) dar (Ohler & Wassarman, 2010). Diese Markierungen dienen als Zeichen für offene und aktive Promoterregionen (Taverna et al., 2007; Suganuma & Workman, 2008; Cler et al., 2009). Die Expressionsstärken breiter Core-Promoter sind allerdings meist geringer als bei Core-Promoter mit TATA-Box oder Inr (Juven-Gershon et al., 2008; Sandelin et al., 2007). Diese Möglichkeiten der TFIID-Rekrutierung bilden folglich Alternativen oder Ergänzungen zu der festen Bindung zwischen TBP, TAFs und Promoter-DNA, die z.B. in CpG-Islands genutzt werden können.

Entgegen älteren Darstellungen, der Core-Promoter sei lediglich ein simpler Ein-/Ausschalter der Transkription, ist heute bekannt, daß eine Vielzahl, in Funktion und Struktur unterschiedlicher, Core-Promotermodelle existieren. Da die Core-Elemente häufig nicht den optimalen Konsensussequenzen entsprechen, können Kombinationen aus zwei oder mehr Core-Promoterelementen die PIC-Rekrutierung modulieren und tragen somit auch zur Regulation der betreffenden Gene bei (Smale, 2001). Neben der Sequenzqualität eines Core-Promotermotives ist auch die Distanz zwischen den Motiven ein kritischer Faktor. Wie von Juven-Gershon et al. (2008) beschrieben, können bereits kleine Distanzunterschiede von nur wenigen Nukleotiden zwischen zwei kombinierten Core-Promoterelementen, die Expression drastisch reduzieren.
Daß der Core-Promoter eine wichtige Komponente in der Genexpression ist, wird auch durch die Anzahl alternativer und gewebespezifischer TBPs und TFIID-Untereinheiten sichtbar. Die *in-vitro* Entwicklung eines „Super Core-Promoters“ (SCP) aus TATA-, Inr-, MTE- und DPE-Motiv zeigt, daß die Modulation der Expressionsstärke nicht nur wie bisher angenommen durch assoziierte TFs oder Enhancer-/Repressor-Elemente erfolgt, sondern auch direkt durch die Struktur des Core-Promoters bestimmt wird (Juven-Gershon et al., 2006b).

1.7.2. Alternative Core-Promoter-Erkennungsfaktoren

Neben den oben beschriebenen basalen TFs sind in höheren Eukaryoten zusätzlich viele Gene bekannt, die für TRFs (TBP-related factors) und alternative TAFs kodieren. Es wird davon ausgegangen, daß diese alternativen Untereinheiten die Promoterspezifität der somit unterschiedlichen TFIID-Varianten beeinflussen und die Diversität der RNAPII Core-Promoter weiter erhöhen (DeJong, 2006; Sikorski & Buratowski, 2009; Müller et al., 2010). Der Fokus der hier vorgestellten alternativen TRFs und TAFs liegt auf Faktoren mit Keimbahnfunktionalität.

In Eukaryoten existieren insgesamt drei TRFs, die TRF1, TRF2 und TRF3 genannt werden. TRF1 kommt nicht in Human und Hefe vor und ersetzt in *Drosophila* das TBP in RNAPIII-abhängiger Transkription (Juven-Gershon & Kadonaga, 2010).
Das TRF2-Protein (alternative Namen: TRF, TLP, TLF, TRP) ist aus verschiedenen Spezies bekannt, aber zwischen den Spezies weniger konserviert als TBP. Obwohl in Human die TRF2 Core-Domäne, die für die TATA-Box Erkennung verantwortlich ist, zu 60% mit TBP übereinstimmt, ist TRF2 nicht in der Lage selbstständig an die TATA-Sequenz zu binden. TRF2 assoziiert hierfür mit TFIIA und TFIIB. In männlichen Keimzellen komplexieren sowohl TRF2 als auch TBP mit einem TFIIA paralogen Protein ALF, „TFIIA-like factor“ (Reina & Hernandez, 2007). In Säugern ist die höchste TRF2-Expression in Keimzellen zu finden. *Trf2* Knock-out Mäuse weisen Schäden in der Spermatogenese auf (Goodrich & Tjian, 2010). Einige Daten

weisen auch darauf hin, daß TRF2 die Transkription von TATA-losen Core-Promoterregionen begünstigt und Promoter mit TATA-Box, über Sequestrierung von TFIIA, reprimieren kann. Für das *Histon1* Gen konnte in *Drosophila* eine Verbindung zwischen TRF2, TATA-losem Promoter und einer Pyrimidin-reichen TC-Box Region im Promoter beschrieben werden, die ausschließlich von TRF2 und nicht von TBP erkannt wird. *trf2* Knock-out Fliegen sind nicht lebensfähig und eine partielle TRF2-Inaktivierung führt unter anderem zu Fehlern in der Differenzierung von männlichen und weiblichen Keimzellen (Kopytova et al., 2006; Shima et al., 2007). Es scheint also möglich, daß die Keimzellentwicklung durch atypische TRFs und TAFs, als Alternative zum Core-Promoter-Erkennungsfaktoren TFIID, reguliert wird (Goodrich & Tjian, 2010).

TRF3 (alternative Namen: TBP2, TBPL2) welches wie TBP die TATA-Box erkennt, konnte in Mensch, Maus und Xenopus nachgewiesen werden, und scheint nur in Vertebraten vorzukommen (Persengiev et al., 2003; Reina & Hernandez, 2007). In terminal differenzierten Myoblasten scheint der TFIID-Komplex durch einen alternativen TRF3–TAF3-Komplex ersetzt zu werden (Deato & Tjian, 2007, 2008; Deato et al., 2008). Eine TRF3 Keimbahnfunktion ist für Maus Oozyten bekannt und scheint TBP während der Follikulogenese komplett zu ersetzen (Goodrich & Tjian, 2010). *Trf3* Knock-out Mäuse sind lebensfähig, aber infertil.
Neben den alternativen TRFs mit Keimbahnfunktion, konnten ebenso TAFs mit keimzellspezifischer Funktion für mehrere Spezies beschrieben werden (Goodrich & Tjian, 2010; Müller et al., 2010). Im Drosophilagenom sind fünf zusätzliche Testis (t)TAF-Paraloge kodiert, die während der Spermatogenese exprimiert werden (Hiller et al., 2001, 2004). Diese alternativen tTafs sind in Säugetieren allerdings nicht konserviert.

Dagegen wird in Maus Spermatogonien und frühen Spermatozyten ein TAF7-Paralog, TAF7L genannt, exprimiert, welches postmeiotisch stark in den Kern importiert wird (Pointud et al., 2003). Ebenfalls in Maus ist TAF4b mit Expression in weiblichen Granulosazellen und männlichen Spermatiden bekannt (Goodrich & Tjian, 2010). In adulten Mäusen scheint TAF4b für den Erhalt von Keimzellen und Hoden eine Rolle zu spielen. Wiederum in Altweltaffen, den Catarrhini, ist zu dem X-chromosomalen TAF1 ein aktives Retrotransposon TAF1L bekannt, welches die meiotische X-Inaktivierung durch Lokalisation auf Chromosom 9 umgeht (Wang & Page, 2002). Bestimmte TAFs, wie z.B. TAF9, können neben der DNA-Bindung auch mit bestimmten TFs interagieren und hierüber Gewebespezifitäten erzeugen (Sengupta et al., 2009).
Zusammengenommen scheinen also in männlichen, wie weiblichen Keimzellen spezifische Transkriptionsinitiationskomplexe die kanonischen TFIID- und TFIIA-Komplexe im PIC funktionell zu ersetzen (Müller et al., 2010). D.h. also, daß nicht nur die diversen Strukturen des

Core-Promoters, sondern auch eine Vielzahl alternativer Core-Promoter-Erkennungsfaktoren zu einer, unabhängig von zusätzlichen Enhancer- und Repressorelementen, hohen Plastizität der Transkriptionsmaschinerie und somit der Genregulation führen (Smale, 2001; Ohler & Wassarman, 2010).

1.7.3. Promoterstruktur, Enhancer, Mediator-Komplex und DNA-Signaturen

Die eukaryotische Transkriptionsregulation erfordert die Integration einer Vielzahl komplexer Informationen, die als Signale im Core-Promoter, einem Nexus gleich, zusammenlaufen (Heintzman & Ren, 2007). Die Regulation der Genexpression ist in einem hochkomplexen „regulativen Code“ niedergelegt, der aus bestimmten „*cis*-regulativen Sequenzen“ (Promoter, Enhancer, Silencer, Insulator, Matrix/Scaffold Attachment Region), allgemeinen und spezifischen *trans*-Faktoren (GTFs, TFs, Chromatin-bindende Proteine) und epigenetischen Markierungen besteht, die im räumlichen und zeitlichen Kontext des dynamisch veränderbaren Chromatins koordiniert werden müssen (Wray et al., 2003; Heintzman & Ren, 2007). Vereinfachte Darstellungen dieser komplexen Prozesse sind schematisch in den Abbildungen 1-11 bis 1-13 dargestellt.

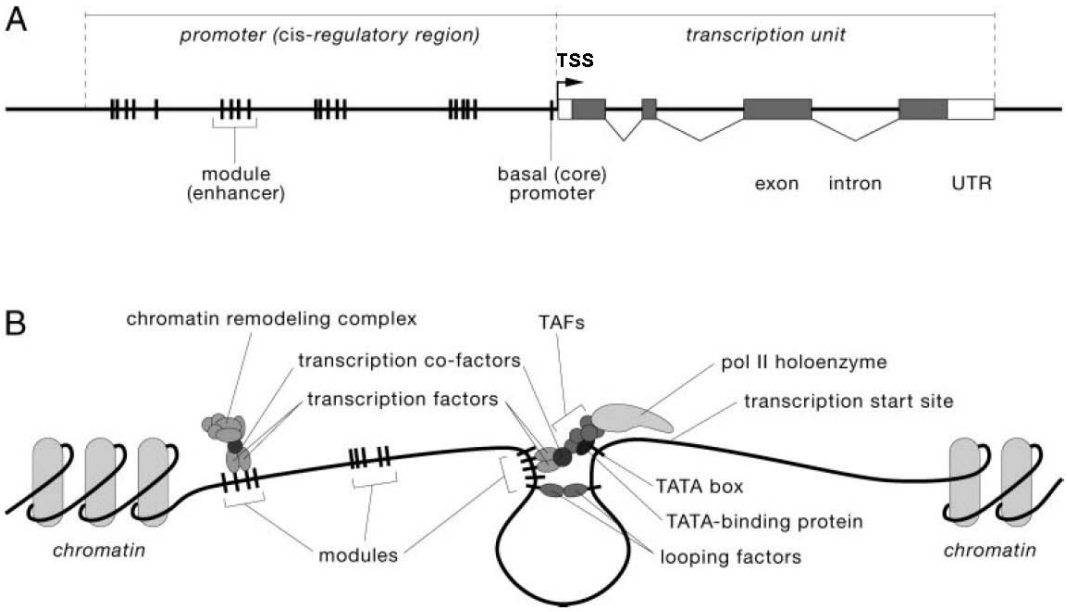


Abb. 1-11: Eukaryotische Promoterstruktur und Promoterfunktion.
A: Vereinfachte schematische Darstellung der Eukaryoten Promoterorganisation. Die TSS ist durch einen Pfeil angegeben. Proximale und distale Promoterelemente sind meist innerhalb einiger kb stromaufwärts der TSS lokalisiert. Entfernte regulative Elemente, die bis zu Megabasen entfernt sein können, sind nicht gezeigt. Horizontale Striche geben TF-Bindestellen an, die funktionell als TF-Module und Enhancer agieren können. Exone sind durch dunkle Boxen, UTRs (untranslated region) durch helle Boxen markiert. Zu der oben gezeigten idealisierten Promoterorganisation sind viele weitere Konfigurationen möglich.
B: Schematische Darstellung eines aktiven Eukaryotenpromoters. (Verändert nach Wray et al., 2003).

Die wichtigste Voraussetzung für die Charakterisierung einer Promoterregion ist die genaue Eingrenzung der TSS, da diese üblicherweise 5´ der Transkriptionseinheit lokalisiert ist. Der Core-Promoter erstreckt sich über ~ 40 Nukleotide stromauf- und abwärts der TSS (Kapitel 1.7.1) und der proximale Promoter reicht je nach Definition bis ~ 250 bp stromaufwärts der TSS (Abb. 1-11; Wray et al., 2003; Heintzman & Ren, 2007). Im proximalen Promoter sind ebenfalls *cis*-regulative Sequenzen vorhanden, die sich aber von denen des Core-Promoters unterscheiden und z.B. gewebespezifische TF-Bindestellen enthalten können. Core-Promoter und proximaler Promoter werden meist einfach gemeinsam als Promoter bezeichnet (Heintzman & Ren, 2007). Eine weitere Ebene der Genregulation entsteht durch die Nutzung alternativer Promoterregionen und damit verbunden der Transkriptstart an alternativen TSS. Schätzungen gehen davon aus, daß ~ 52% der annotierten Humangene über „potentielle alternative Promoter“ (PAP) reguliert werden können (Kimura et al., 2006). Im Mittel gehören 3,1 PAPs, mit einem Anteil von 17% gewebespezifisch genutzter PAPs, zu den jeweiligen Transkriptionseinheiten (Kimura et al., 2006). Im Unterschied zu einem breiten Core-Promoter mit flexibler TSS-Position, sind alternative TSSs deutlich distanziert und sind häufig unter veränderten Bedingungen (Gewebe, Geschlecht, Entwicklungszeitpunkt) aktiv (Ohler & Wassarman, 2010).

Darüber hinaus können zusätzliche funktionelle „*cis*-regulative Elemente“, auch „*cis*-regulative Module“ genannt (CRM), stromaufwärts und stromabwärts einer Transkriptionseinheit, im Intron der betreffenden Transkriptionseinheit, sowie in Introns benachbarter Gene, liegen (Wray et al., 2003; Davidson, 2006). CRMs sind als DNA-Sequenzen definiert, die üblicherweise 100 bis 1000 bp lang sind, viele modular organisierte TFBSs beinhalten und die Expression assoziierter Gene regulieren können (Davidson, 2006; Jeziorska et al., 2009). CRM wird als Überbegriff für *cis*-regulative Elemente, wie z.B. Promoter, Enhancer oder Silencer, verwendet (Jeziorska et al., 2009). Enhancer sind positiv wirkende CRMs, die für eine korrekte räumliche, zeitliche und quantitative Expression eukaryotischer Genen notwendig sind, da der Core-Promoter, trotz vieler spezifischer Signale, häufig nur die Aktivierung der Transkription, nicht aber die hochkomplexe Zelltyp-spezifische Feinsteuerung bewirken kann (Kleinjan & van Heyningen, 2005; Wray et al., 2003; Smale & Kadonaga, 2003; Bulger & Groudine, 2010). Bisher beschriebene Enhancer können bis zu 1,5 Mb von der dazugehörigen Transkriptionseinheit entfernt sein (Smale & Kadonaga, 2003; Wray et al., 2003; Kleinjan & van Heyningen, 2005; Rauschendorf, 2006; Epstein, 2009). Enhancer z.B. sind typischerweise zwischen 100 und 500 bp lang und beinhalten Cluster von TF-Bindestellen (TFBS), die wiederum etwa 5 bis 12 bp lang und degeneriert sind (Smale & Kadonaga, 2003; Wray et al., 2003). D.h. für jeden TF gibt es eine Vielzahl von möglichen Sequenzen an die er, mit unterschiedlicher Affinität, binden kann. Die TFBS sind meist in Form von TF-Modulen, Einheiten synergistisch agierender TFs, organisiert. Wie viele TFBS in einem Enhancer vorhanden sein müssen, bzw. wie viele TFs synergistisch

zusammenarbeiten müssen, um einen funktionellen Enhancer zu etablieren, ist unterschiedlich. Eine Definition geht von einem Minimum von zehn TFBS für mindestens drei verschiedene TFs aus (Levine & Tjian, 2003). Davidson (2006) erwähnt in seiner CRM-Definition dutzende TFBSs für mindestens drei bis zehn unterschiedliche Transkriptions-faktoren, die in TF-Familien, anhand ihrer gemeinsamen DNA-binde Domänen, gruppiert werden. In der Systembiologie wird versucht die charakteristischen „Gen-regulativen Funktionen" der verschiedenen CRM-Typen zu beschreiben, indem CRM-Qualitäten, wie z.B. der quantitative Besatz mit TFs (Input), mit der Promoteraktivität (Output) in Bezug zueinander gesetzt werden (Teif, 2010).

Klassische Methoden zur Identifizierung solcher Enhancer-Elemente waren bisher das Screening von Mutationen mit pathologischem Phänotypen, trotz intakter Transkriptions-einheiten, oder aufwendige Genomvergleiche evolutionär entfernter Spezies („phylogenetischer Sequenzvergleich"), zur Lokalisation von „conserved noncoding sequences" (CNS; Hardison, 2000; Loots et al., 2000; Elgar & Vavouri, 2008; Loots, 2008). Vertebratengenome enthalten hochkonservierte single-copy Sequenzen, die nicht transkribiert werden, und beim Menschen 1 bis 3% des Genoms ausmachen (Dermitzakis et al., 2005; Lettice und Hill, 2005). Diese CNSs sind von bisher unbekannter Funktion, aber es ist anzunehmen, dass sich in diesen Bereichen vielfach *cis*-regulative Elemente, wie z.B. Enhancer, befinden.

Mit modernen genomweiten Microarray-Studien und Hochdurchsatz-Sequenzierungen (chIP-Seq) konnten aktive regulative Enhancer-Elemente auch durch eine bestimmte Chromatin-signatur (Monomethylierung von H3K4me1) in Assoziation mit bestimmten TFBSs für den Faktor p300 („Histone Acetyltransferase p300") identifiziert werden (Heintzman et al., 2007; Bulger & Groudine, 2010). Im Unterschied dazu werden aktive Promoterregionen durch Trimethylierung von H3K4 (H3K4me3) und die korrekte Elongationsrichtung der RNAPII, 5´ → 3´, wird ebenfalls durch Chromatinmarkierungen, H3K79me2 und einem Gradienten von H3K4me3 zu H3K4me1 am Ende der Transkriptionseinheit, markiert (Seila et al., 2008; Heintzman et al., 2007; Li et al., 2007). Eine solche Studie fand 55000 Sequenzen, die dieser Enhancer-Chromatinsignatur entsprachen (Heintzman et al., 2009). Rein rechnerisch, sollten also auf jedes Humangen mehrere *cis*-regulative Elemente kommen, die jeweils die Genregulation spezifisch modulieren können (Ben-Tabou de-Leon & Davidson, 2007; Bulger & Groudine, 2010; Nam et al., 2010). Umgekehrt gilt aber auch, daß ein *cis*-regulatives Element mehrere Gene regulieren kann (Teif, 2010).

Möglich wird die Integration entfernter Enhancersignale durch die dreidimensionale Organisation der Chromosomen im Zellkern. Die Chromosomen besetzen bestimmte Regionen im Zellkern, Chromosomen-Territorien genannt, die für aktives und inaktives Chromatin, sowie nach Replikation oder Transkription der verschiedenen Gene, organisiert sind (Cremer & Cremer, 2001; Ohlsson et al., 2010). Diese genomweite Chromatinorganisation kann Enhancer in räumliche Nähe zu einem Promoter in einer aktiven Chromatindomäne bringen und damit

erst eine korrekte Genexpression ermöglichen (Abb. 1-12; Lieberman-Aiden et al., 2009; Bulger & Groudine, 2010; Raab & Kamakaka, 2010).

Entscheidend für die Etablierung unterschiedlich regulierter Chromatindomänen sind die sogenannten S/MARs („Scaffold/Matrix Attachment Regions“), DNA-Abschnitte, die eine Organisation des Chromatins in Schleifen ermöglichen. Die S/MARs wurden von zwei Forschungsgruppen, durch verschiedene Methoden unabhängig voneinander identifiziert und einerseits als SARs (Gasser & Laemmli, 1986a, 1986b) und andererseits als MARs (Cockerill & Garrard, 1986a, 1986b) beschrieben. Nachdem sich herausgestellt hatte, daß es die gleichen DNA-Abschnitte waren, ist der Begriff S/MAR für diese (A,T)-reichen Kernmatrix-assoziierten DNA-Sequenzen eingeführt worden.

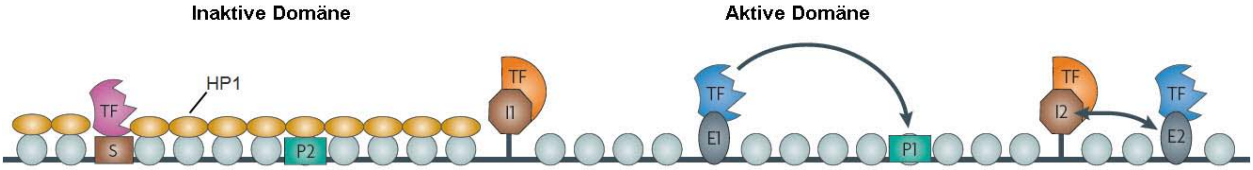


Abb. 1-12: Chromatindomänen und regulative Elemente im Schema.
Die Chromatindomänen sind linear dargestellt. Tatsächlich ist die Chromatinorganisation dreidimensional im Zellkern etabliert, wodurch z.B. die Interaktion entfernter DNA-Bereiche, wie zwischen Enhancer (E1) und Promoter (P1), erst ermöglich wird. Das Silencer-Element der inaktiven Promoterdomäne initiiert mit spezifischen TFs eine Heterochromatinbildung, durch z.B. „Heterochromatin Protein 1“ (HP1). Das Insulator-Element 1 (I1) verhindert eine Ausbreitung des Heterochromatins in die aktive Chromatindomäne, die durch I1 und I2 begrenzt wird. Das Enhancerelement E1 wird durch spezifische TFs gebunden und kann die Promoterfunktion (P1) modulieren. Ein zweites Enhancerelement (E2) kann nicht mit dem Promoter P1 kommunizieren, da es außerhalb der aktiven Domäne lokalisiert ist und die Signale durch I2 blockiert werden. Hellblaue Kreise stellen Nukleosomen, gelbe ovale stellen repressive Proteine dar. S = Silencer, P = Promoter, I = Insulator, E = Enhancer, TF = Transkriptionsfaktor. (Verändert nach Raab & Kamakaka, 2010).

Die 30nm-Chromatinfasern sind in ~ 5 bis 200 kb große, voneinander unabhängige Schleifen („loop domains“) an der inneren Kernhülle, der Kernmatrix, befestigt (Paulson & Laemmli, 1977; Beggs & Migeon, 1989; Chernov et al., 2004; Dillon, 2006). Im Interphasekern dienen die S/MARs dieser Chromatinschleifen zur Etablierung und Trennung aktiver und inaktiver chromosomaler Regionen (Heng et al., 2004; Platts et al., 2006; Linnemann et al., 2009; Wang et al., 2010). Der Vorgang der Schleifenbildung ist dynamisch und wird während des Zellzykluses für jeden Zelltyp spezifisch angepaßt (McKnight et al., 1992; Bonifer et al., 1994). S/MARs sind häufig an den Chromatin-Domänengrenzen kodierender DNA-Bereiche lokalisiert, so daß das Chromatin in unabhängige Bereiche unterteilt und vor externen Einflüssen abgeschirmt werden kann (Cockerill et al., 1987; Jarman & Higgs, 1988; Namciu et al., 1998; Mielke et al., 2002). S/MARs wird daher auch eine Art Isolatorfunktion zugeschrieben, die z.B. durch Bindung des „MAR-binding Protein“ (MARBP) CTCF („CCCTC-binding factor“) erfolgt (West et al., 2002; Wallace & Felsenfeld, 2007; Ohlsson et al., 2010; Raab & Kamakaka, 2010). S/MAR-Lokalisationen in Intronsequenzen von Transkriptionseinheiten, mit Einfluß auf die Gen-

expression, wurden ebenfalls beschrieben (Cockerill et al., 1987; Rudd et al., 2004). Schübeler et al. (1996) berichten von kontext- und entfernungsabhängigen Effekten auf die Transkription. Eine 800 bp lange S/MAR Sequenz wurde in unterschiedlichen Entfernungen zu einer TSS in einen Vektor für stabile Transfektion kloniert. In einem Abstand von 4 und mehr kb konnte eine aktivierende Wirkung auf die RNA-Expression festgestellt werden. Im Abstand von weniger als 2,5 kb wurde eine fast vollständige Unterdrückung der Expression festgestellt (Schübeler et al., 1996). Das Ergebnis der stabilen Transfektion war die Aufrechterhaltung einer stabilen Langzeitgenexpression.

Es wird auch immer deutlicher, daß nicht alle S/MARs dieselben Funktionen im Zellkern besitzen. Die unterschiedlichen Funktionen sind zum Teil mit unterschiedlichen MARBPs und deren Expressionsprofilen zu erklären. MARBPs können so im Zusammenspiel mit aktivierenden oder reprimierenden Co-Faktoren direkt Einfluß auf die Gen-expression nehmen. Dies erfolgt über die Einleitung von Chromatinmodifikationen oder im Komplex mit der Transkriptionsmaschinerie (Nayler et al., 1998; Cai, et al., 2003; Galande et al., 2007; Garee & Oesterreich, 2010; Wang et al., 2010).

Die Integration distaler *cis*-regulativer Aktivator- oder Repressorsignale auf den Core-Promoter bzw. den PIC erfolgt schließlich über einen bis zu 1 MDa großen und aus bis zu 30 Untereinheiten bestehenden Multiprotein-Komplex (Blazek et al., 2005; Fuxreiter et al., 2008; Tóth-Petróczy et al., 2008; Malik & Roeder, 2010).

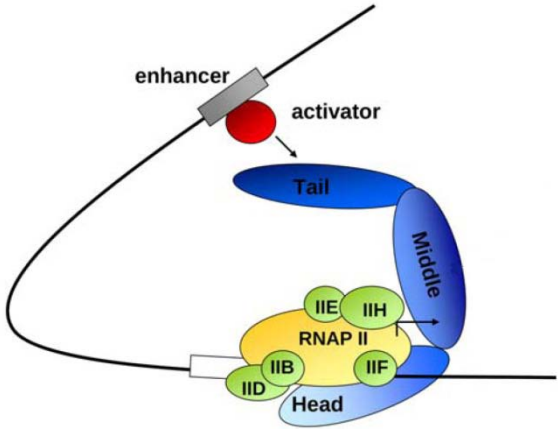


Abb. 1-13: Übertragung entfernter genspezifischer TF-Signale auf die Transkriptionsmaschinerie durch den verformbaren Mediator-Komplex.

Der Mediator ist hier dreiteilig als Head-Middle-Tail dargestellt. Der als „Tail" bezeichnete Komplex-Schwanz interagiert mit einer Reihe an Aktivator- und Repressor-TFs. Die Signalübertragung, auf den mit RNAPII in physischem Kontakt stehenden Kopfteil, erfolgt über das Mittelstück (nach Tóth-Petróczy et al., 2008).

Dieser wird als Mediator-Komplex bezeichnet und stellt den am häufigsten vorkommenden Co-Aktivator-Komplex dar (Abb. 1-13). Der Mediator ist in der Lage TFs auf der einen Seite und den RNAPII-Komplex auf der anderen Seite zu binden (Abb. 1-13), und durch Konformations-änderungen im Mediator, Signale auf den PIC zu übertragen. Zusätzlich zu diesem Schwanz-zu-Kopf Informationstransfer können zahlreiche Mediator-Untereinheiten mit TFs interagieren (Malik & Roeder, 2010).

Spezifische DNA-Signaturen im Bereich von Promotersequenzen

Da die DNA im Chromatin eingebettet ist, muß die DNA in der Core-Promoterregion für den PIC und für TFs, durch Öffnung der Chromatin-Verpackung, zugänglich gemacht werden (Heintzman & Ren, 2007). Neben den bereits beschriebenen Histonmodifikationen, die die Zugänglichkeit zur DNA beeinflußen, nimmt auch die Basenzusammensetzung eines DNA-Stranges, durch Struktur, Windung und Krümmung, starken Einfluß auf Organisation und Besatz der DNA mit Histonen in den Nukleosomen.

Je nach Promoterarchitektur und Expressionszustand kann die Region direkt stromaufwärts der TSS konstitutiv zugänglich sein, oder nach einem aktiven Prozeß des „Chromatin-Remodellings“ geöffnet werden. Für eine aktive Transkription ist die Region bis ~ 150 bp stromaufwärts der TSS Nukleosomen-frei und wird „nucleosome-depleted region“ (NDR) genannt. Die Länge einer NDR entspricht meist ziemlich exakt der Länge einer DNA-Doppelhelix mit 1,65 Windungen (147 bp), welche um ein Histon-Oktamer gewunden ist. Der mittlere Abstand zwischen zwei Nukleosomenmitten beträgt für Humanchromatin 185 bp wovon ~ 38 bp auf den DNA-Linker entfallen (Jiang & Pugh, 2009). Bei Metazoen liegt die TSS innerhalb der NDR (Jiang & Pugh, 2009). Im Gegensatz zur NDR sind die Nukleosomen-positionen direkt stromabwärts der TSS (+1 Nukleosom) und direkt stromaufwärts der NDR (-2 Nukleosom)[9] meist sehr genau bestimmt (Schones et al., 2008; Jiang & Pugh, 2009; Radman-Livaja & Rando, 2010). Eine permanent geöffnete NDR wird häufig durch binäre oder homopolymere DNA-Trakte erreicht (Gal et al., 2003; Bajic et al., 2006; Yagil, 2006; Cairns, 2009; Segal & Widom, 2009a, 2009b). Von solchen Binären W[10]- und S[11]-reichen Trakten ist bekannt, daß sie Einfluß auf die DNA-Windung und auf die Trennung der komplementären Stränge ausüben, was letztendlich notwendig für Replikation und Transkription ist. W-Sequenz-abschnitte schmelzen schneller und einfacher, als andere Basenzusammensetzungen.
Poly(dA.dT)[12]-reichen Sequenzen, mit 10 bis 20 bps Länge, die die normale Windung der B-Form-DNA verhindern und somit der Komplexbildung des Histon-Oktamers entgegenwirken, werden „nucleosome excluding sequences“ (NES) genannt (Schones et al., 2008; Cairns, 2009; Jiang & Pugh, 2009; Segal & Widom, 2009a; Radman-Livaja & Rando, 2010). Vergleichbar zu den NES existieren auch Nukleotidsignaturen, die den Einbau von Nukleosomen begünstigen. Die „nucleosome positioning sequences“ (NPS) bestehen meist aus AA/TT und CC/GG Dinukleotiden mit einem Intervall von 10 nts, was die DNA-Windung, um ein Histon-Oktamer, erleichtert.

---

9 Nomenklatur der Human-Nukleosomen nach Scholes et al. (2008).
10 W = A,T.
11 S = C,G.
12 poly(dA.dT) = „Homopurin.Homopyrimidine-Trakt “ oder „homopolymeric Tracts“ (HT).

Als weitere Signatur treten ebenfalls R.Y-Trakte[13] gehäuft in Promoterregionen auf. Für Polypyrimidin-reiche Sequenzen in Promoterregionen konnte in mehreren Fällen ein positiver Einfluß auf die Transkription in Human-, Maus-, und Drosophila-Genen gezeigt werden (McKeon et al., 1984; Hoffman et al., 1990; Cameron et al., 1992; DiMartino et al., 1994; Dörflinger et al., 1999; Holmes & Tjian, 2000; Timón & Beverley, 2001; Smale & Kadonaga, 2003; Miyahara et al., 2004). In Pflanzen konnte gezeigt werden, daß TC-Elemente in AT-armen Promoterregionen eine neue Klasse von Core-Promotermotive darstellen können (Bernard et al., 2010). K.M-DNA-Trakte[14] kommen ebenfalls gehäuft in Promoterregionen vor, sind allerdings bisher funktionell kaum beschrieben (Gal et al., 2003; Yagil, 2006).

Das S-Motiv, mit dem Dinukleotid 5´-CpG-3´, ist vor allem als CpG-Island (CGI) bekannt. Im Humangenom sind CpG-Dinukleotide nicht gleichmäßig über das Genom verteilt, sondern kommen in CGI-assoziierten Promoterregionen bis zu 20mal häufiger vor als in anderen Genomregionen. Insgesamt kommen nur etwa 10 bis 20% der erwarteten CpG-Dinukleotide in der DNA des Humangenoms. Ursache ist, daß außerhalb der Promoterregionen das Cytosin der CpG-Dinukleotide meist methyliert vorliegt. Durch Desaminierung dieses 5-Methyl-Cytosins entsteht in der Folge das Nukleotid Thymin. Somit wird der genomische Anteil von CpG-Dinukleotiden zu Gunsten von TpG-Dinukleotiden reduziert (Knippers, 2001).

CpG-Dinukleotide entstehen aber auch durch Punktmutationen neu, so daß ein konstanter Wert von etwa 20% der erwarteten CpG-Dinukleotid Frequenz im Genom erhalten bleibt. Der durchschnittliche (C,G)-Gehalt des Humangenoms beträgt 41%. Der durchschnittliche (C,G)-Gehalt eines CGIs beträgt dagegen 67%. Ein Kriterium für CGIs ist das Auftreten von CpG-Dinukleotiden mit einer Häufigkeit, die über 60% des erwarteten Wertes, bezogen auf den (C,G)-Gehalt des CGIs, liegt. Bei 200 bp macht dies in etwa 8 oder mehr CpG-Dinukleotide aus (Gardiner-Garden & Frommer, 1987; Rozenberg et al., 2008). In CGIs kommen CpG-Dinukleotide mit der normalen, erwarteten Frequenz vor. Die Länge eines CGIs ist, für eine potentielle Promoterfunktion, ein wichtiges Kriterium und beträgt je nach Definition mindestens 200 bis 500 bp (Gardiner-Garden & Frommer, 1987; Takai & Jones, 2002).

Kürzere CpG-Abschnitte, CpG-Islets (CpG-Inselchen) genannt, sind ebenfalls häufig unmethyliert und zwischen Spezies konserviert. Sie sind unter 200 bp lang und können mit alternativen TSSs assoziiert sein (Hackenberg et al., 2006, 2010).

NDRs können auch durch ein CpG-Island erzeugt werden, die ebenfalls zu einer lokalen Nukleosomeninstabilität im Bereich der TSS führen (Li et al., 2007; Tora & Timmers, 2010). CGIs können auch mit „Z-DNA forming regions“ (ZDRs), Regionen links gewundener DNA, zusammenfallen, die ebenfalls als Sequenzbesonderheit stromaufwärts der TSSs vorkommen (Rich et al., 1984; Champ et al., 2004). CGI und ZDR bedingen sich allerdings nicht gegenseitig

[13] R.Y-Trakt = „Polypurin.Polypyrimidin“-Trakt.
[14] K = G,T und M = A,C.

und kommen auch unabhängig voneinander vor. Tazi und Bird (1990) beschreiben, daß „CpG-Island Chromatin“, im Gegensatz zum „Gesamt-Chromatin“, wesentlich weniger Histon H1 beinhaltet, die Histone H3 und H4 verstärkt acetyliert sind, was ein Zeichen für aktives Chromatin ist. Neuere Daten zeigen, daß das „CpG-binding Protein“ (CXXC1, CFP1) an unmethylierte CGI-DNA bindet und zusammen mit dem SET1 Histon-Methyltransferase Komplex für die Methylierung von Histon 3 Lysin 4 (H3K4) verantwortlich ist (Dehé et al., 2006; Thomson et al., 2010).

50 bis 70 % aller Humangene sind mit CGIs assoziiert, die in deren Promoterregionen vorkommen (Cross & Bird, 1995; Antequera, 2003; Sandelin et al., 2007; Rozenberg et al., 2008; Suzuki & Bird, 2008; Zhao & Han, 2009; Hackenberg et al., 2006, 2010). CGIs haben für die Genregulation eine wichtige Rolle, wenn durch epigenetische Modifikationen die Cytosin-Reste methyliert werden, was mit einer Inaktivierung der entsprechenden Promoterregion einhergeht. Entwicklungs- und gewebespezifische Methylierungsmuster in CGIs haben somit einen großen Einfluß auf die Regulation spezifisch exprimierter Gene.

1.8. Post-transkriptionelle Regulation der Genexpression

Die post-transkriptionelle Kontrolle der Genexpression umfaßt viele Regulationsebenen wie mRNA-Export, mRNA-Lokalisation, mRNA-Speicherung, mRNA-Stabilität, mRNA-Translation und mRNA-Turnover (Mata et al., 2005). Die meisten dieser Schritte, die letztlich die translatierte Proteinmenge bestimmen, erfolgen im Zytoplasma der eukaryotischen Zelle und sind untereinander vernetzt (Mata et al., 2005). Der limitierende Faktor der Proteintranslation ist hauptsächlich die Effektivität der Translationsinitiation, die den Besatz einer mRNA mit Ribosomen kontrolliert, und die nachfolgende Bildung des 43S[15]-Präinitiationskomplexes[16] regelt (Mata et al., 2005; Pickering & Willis, 2005; Jackson et al., 2010). Je dichter die mRNA mit Ribosomen belegt ist, desto mehr Protein kann synthetisiert werden. Ein Ribosom belegt ~ 30 Nukleotide auf der mRNA. Die maximale Ribosomen-Dichte wird selten erreicht, da dies meist durch den Initiationsschritt limitiert wird (Mata et al., 2005). Die Transkriptlänge ist negativ zur Ribosomen-Dichte korreliert. Die Ribosomen-Dichte nimmt mit zunehmender Transkriptlänge ab, was wiederum mit einer weniger effizienten Translationsinitiation langer Transkripte zusammenzuhängen scheint (Mata et al., 2005).
Die Translationsinitiation erfordert meist die Bindung der terminalen mRNA 5´-Cap-Struktur (ein 7-Methylguanosin) durch den eIF4F-Multiproteinkomplex (eukaryotic translation initiation factor).

---

15 S = Sedimentationskoeffizent in Svedberg-Einheiten (≙ 10^{-13} sec).

16 43S Präinitiationskomplex besteht aus: 40S-Ribosom + eIF1 + eIF1A + eIF3 + GTP-abhängiges eIF2 + Methionin-beladene Initiator-tRNA (Met-$tRNA_i^{Met}$).

Die Erkennung der Cap-Struktur erfolgt durch die eIF4F-Untereinheit eIF4E. Die Translationseffizienz von mRNAs mit strukturierter, komplexer 5´UTR kann durch eine Überexpression von eIF4E gesteigert werden (von der Haar et al., 2004). Zu dem eIF4F-Komplex gehören die DEAD-Box RNA-Helikase eIF4A und das Multidomänenprotein eIF4G (Filipowicz et al., 2008; Jackson et al., 2010). Die Funktion von eIF4G ist die Bildung eines RNA-Ringes zwischen 5´ und 3´ Ende der mRNA. Erst dieser Ringschluß zwischen eIF4G und dem Poly-A-Bindeprotein PABP ermöglicht den Aufbau des Translationsinitiationskomplexes und somit eine effiziente Translation (Abb. 1-14). Zusammen mit einem weiteren Multiprotein-Faktor eIF3, wird die kleine ribosomale Untereinheit (40S) an die mRNA rekrutiert, welche dann in der Folge in 5´→3´ Richtung entlang der 5´UTR bis zu einem AUG Translationsstartcodon wandert (5´Cap-abhängiger Scanning-Mechanismus). Zusammen mit eIF4A, eIF4B und eIF4F wird der 43S-Präinitiationskomplex als 48S-Translationsinitiationskomplex bezeichnet (Pisarev et al., 2006). Nach Verbindung mit der großen Ribosomenuntereinheit (60S) zum 80S-Ribosom, beginnt am AUG-Codon die Translation (Wilkie et al., 2003; Hughes, 2006; Filipowicz et al., 2008; Smith, 2008; Groppo & Richter, 2009).

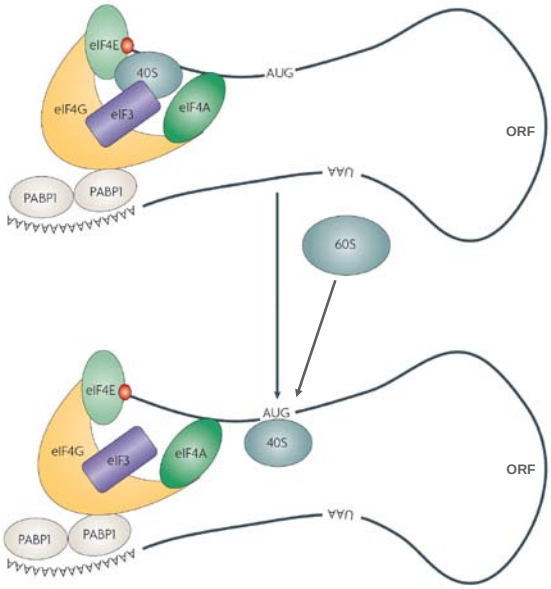


Abb. 1-14: Vereinfachte, schematische Darstellung der Translationsinitiation und der Ringbildung zwischen 5´UTR und 3´UTR.
Der Initiationskomplex besteht aus mindestens zehn unterschiedlichen eIFs, die wiederum selbst Multiproteinkomplexe sein können. Angedeutet ist der Scanning-Mechanismus der 40S-Untereinheit bis zum Antreffen eines AUG Codons („first-AUG rule"). Der rote Punkt stellt die 5´-Cap-Struktur dar. AUG = Translationsstartcodon, UAA = Translationsstopcodon, ORF = Open Reading Frame, eIF = „eukaryotic Translation Initiation Factor", 40S und 60S = ribosomale Untereinheiten, PABP = Poly-A binding Protein. eIF4F = eIF4A + eIF4G + eIF4E. Weitere Erläuterungen siehe Text (nach Filipowicz et al., 2008).

Die Translationsinitiation kann durch vielfältige Prozesse kontrolliert werden, die überwiegend von Sequenz und Struktur der untranslatierten 5´ und 3´ Enden, UTR genannt, abhängen (Gebauer & Hentze, 2004). Dies können die Länge der 5´UTR (Human zwischen 10 und 2860 Nukleotide), deren 5´UTR-Sequenz, der (G,C)-Gehalt, potentielle Sekundärstrukturen, wie RNA-Schleifen und G-Quadruplexe (*cis*-Effekte), sowie Bindungen von Proteinen an die mRNA (*trans*-Effekte), sein. Eine *cis*-basierte 5´UTR-Translationskontrollmöglichkeiten ist das Vorhandensein von uATGs (upstream ATG Translationsstartcodon), die eine Verlängerung des Haupt-Leserahmens im Falle von in-frame uATGs und eine Verschiebung des funktionellen

Leserahmens im Falle von out-of-frame uATGs zur Folge hat. Ein zweites *cis*-Element sind uORFs (upstream open reading frame), zusätzliche unabhängige Leserahmen in der 5´UTR, die meist für kurze, potentielle Peptide kodieren, und damit den Scanning-Mechanismus der kleinen ribosomalen Untereinheit zum Haupt-ATG behindern oder ganz verhindern können. Generell handelt es sich bei diesen beiden 5´UTR Elementen also um negative Regulatoren der Translation.

Weniger als 50% der Ribosomen, die einen uORF translatiert haben, scannen weiter entlang der 5´UTR, um an nachfolgenden ATGs die Translation zu initiieren (Jackson et al., 2010). Ein solcher uORF, der meist kürzer als 30 Codons ist, sollte für eine mögliche stromabwärts gelegene Reinitiation am funktionellen ATG nicht länger als 35 As sein (Rajkowitsch et al., 2004) und sollte mindestens 50 Nukleotide stromaufwärts des Haupt-ATGs entfernt sein (Kozak 2002; Cvijovic et al., 2007). Die uORF-Sequenz selbst hat in Säuger-Systemen meist nur geringen Einfluß auf die Reinitiationsfähigkeit an einem nachfolgenden ATG (Jackson et al., 2010). Im Gegensatz zu uORFs sind uATGs in 5´UTRs statistisch deutlich unterrepräsentiert (Iacono et al., 2005). uORFs sind in bis zu 50% der humanen 5´UTRs vorhanden, die auch über Speziesgrenzen hinweg konserviert sein können. Die Konserviertheit bezieht sich allerdings meist nicht auf die Ebene der Aminosäurensequenz, was dafür spricht, daß es sich bei diesen uORFs eher um regulative Elemente zur Translationskontrolle handelt und nicht um bioaktive Peptide (Iacono et al., 2005; Neafsey & Galagan, 2007; Calvo et al., 2009; Resch et al., 2009). Oyama et al. (2004, 2007) konnten allerdings zeigen, daß einige kleine Peptide (unter 100 AS) aus kurzen kodierenden ORFs (short ORF = sORF) in 5´UTRs stammen.

Der Start der Proteinsynthese beginnt schließlich am Translationsstartcodon AUG. Der optimale Sequenzkontext für eine effiziente Translationsinitiation wurde von Kozak (1984, 1986, 1987a, 1987b) mit folgender Konsensussequenz beschrieben:

```
-6   -3   +1   +4
 g c c R c c A T G G
```

Das Nukleotid an der +4 Position sollte ein Guanin sein, an der -3 Position sollte sich ein Purin befinden. Diese Nukleotide sind für eine starke Initiation entscheidend. Stimmt eine Translationsstart-Sequenz in Position -3 und +4 mit der Konsensussequenz überein, spricht man von einem starken, oder optimalen Kontext. Ist nur eine der beiden Positionen kanonisch wird der Kontext adäquat, oder suboptimal, wenn keine der beiden Positionen konserviert ist, wird der Kontext schwach genannt. Die Cytosine an -1, -2, -4 und -5 sind nicht so stark konserviert, tragen aber zu der Gesamt-Translationseffizienz einer Kozak-Sequenz bei. Guanin- und Adenin-Nukleotide sind an diesen Positionen ebenfalls möglich, bilden aber einen schwächeren Gesamtkontext aus. Das Guanin an Position -6 ist nicht besonders konserviert, aber es konnte für dasβ -Globin-Gen gezeigt werden, daß eine Transversion von G zu C in

dieser Position den Phänotyp einer Thalassämie, durch eine zu geringe Translation von β-Globin, zur Folge haben kann (De Angioletti et al., 2004). In eukaryotischen mRNAs ist insgesamt häufig ein suboptimaler Kontext zu finden (Kochetov, 2005).

Strukturen in der 5'UTR werden vor allem durch (G,C)-reiche, bzw. G-reiche Sequenzen verursacht. Eine Helix-Konformation vergleichbar der DNA-Helix ist für die mRNA ebenfalls möglich. (G,C)-reiche, selbst-komplementäre Sequenzen tendieren dazu Haarnadelstrukturen auszubilden, die unter dem Begriff „Stem-loop" zusammengefaßt werden (Batey et al., 1999). Kurze Haarnadelschleifen werden auch „Hairpin", oder „Hairpin-loop" genannt. (Batey et al., 1999). Die „Stem-loops" entstehen durch intramolekulare Basenpaarung innerhalb einer einzelsträngigen RNA, indem Basenpaarungen zwischen antiparallelen, meist palindromischen, Sequenzabschnitten Duplexstrukturen ausbilden (Abb. 1-15). Abgeschlossen wird ein solcher Stem meist durch eine ungepaarte Schleife, dem „Hairpin loop". In Hairpin- und Stem-loop Strukturen liegen also sowohl Einzelstrang- als auch Doppelstrangbereiche vor. Verantwortlich für die Stabilität einer solchen Haarnadel ist die Länge der Duplex-Region, sowie die Anzahl und Länge ungepaarter Bereiche. Ungepaarte Bereiche bilden in der Folge kleinere Ausbuchtungen, „Bulge" genannt, wenn sie nur auf einer Seite der Schleife auftreten. Sie werden als „Internal-loop" bezeichnet, wenn sie gleichzeitig auf gegenüberliegenden Seiten auftreten. Pseudoknoten („Pseudoknots"), die aus zwei ineinander geschachtelten Haarnadelstrukturen aufgebaut sind, sowie große Schleifenstrukturen („Loops") sind, wenn sie keine eigene Sekundärstruktur aufweisen, instabil (Batey et al., 1999; Chen, 2008; Holbrock, 2008). Pseudoknoten können, bedingt durch deren Instabilität, von Programmen wie mfold nicht vorhergesagt werden (Zuker et al., 1999; Deogun et al., 2004).

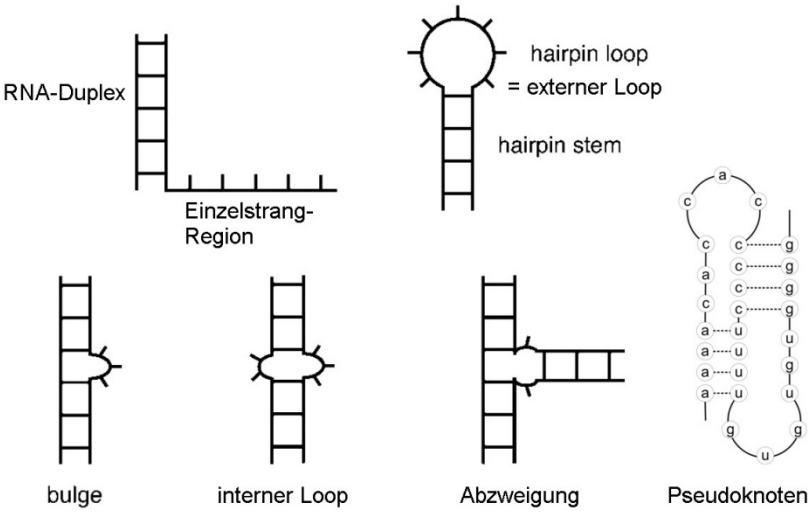


Abb. 1-15: Schematische Darstellung möglicher RNA-Sekundärstrukturen.
Erläuterung siehe Text. (Verändert nach Batey et al., 1999; Krishnan et al., 2010).

Trinukleotid-Repeats des Typs XGG, auch als „Simple Sequence Repeats" SSRs bezeichnet, sind bekannt für die mögliche Ausbildung von stabilen G-Quadruplexstrukturen, die in DNA

sowie RNA vorkommen können (Sha et al., 1999; Zarudnaya et al., 2003; Lawson & Zhang, 2008; Wieland & Hartig, 2007). Mit klassischen Struktur-Vorhersageprogrammen wie mfold können G-Quadruplexe nicht erfasst werden. Die translationsmodulierende Wirkung eines Quadruplexes konnte für das Proto-Onkogen *NRAS* gezeigt werden (Kumari S. et al., 2007, 2008). Potentielle Quadruplexstrukturen, bzw. SSRs liegen vermehrt am 5´Ende von 5´UTRs (Lawson & Zhang, 2008).

RNA-Sekundärstrukturen erschweren also das Scannen der 40S Ribosomenuntereinheit entlang der 5´UTR. Der Scanning-Prozeß kann bereits, ab freien Energien von -50 kcal/mol, in einer Sekundärstruktur, effizient unterbunden werden (Pelletier & Sonenberg, 1985; Davuluri et al., 2000). In Form von IRES (Internal Ribosome Entry Site) und ribosomalen Shunts, nehmen RNA-Sekundärstrukturen aber auch direkten Einfluß auf die Art der Translationsinitiation (Baird et al., 2006, 2007; Hernández, 2008; Morley & Coldwell, 2008; Xia & Holcik, 2009). In beiden Modellen wird der ribosomale Scanning-Mechanismus, in 5´→3´ Richtung entlang der 5´UTR, umgangen und das Ribosom kann direkt zum Haupt-ATG dirigiert werden (Abb. 1-16). RNA-Sekundärstrukturen haben auch fundamentale biologische Bedeutung, wie das Beispiel der Transfer-RNAs (tRNA) zeigt, die während der Proteinsynthese die korrekten Aminosäuren zu den Codons auf der mRNA übermitteln (Kim et al., 1974; Robertus et al., 1974). Zudem können Sekundärstrukturen auch als Ansatzpunkt für mRNA Bindeproteine (mRNPs), wie z.B. für das IRP (Iron-Regulatory Protein), dienen, welches an eine konservierte Struktur namens IRE (Iron Response Element) bindet (Wilkie et al., 2003; Wells, 2006; Chatterjee & Pal, 2009).

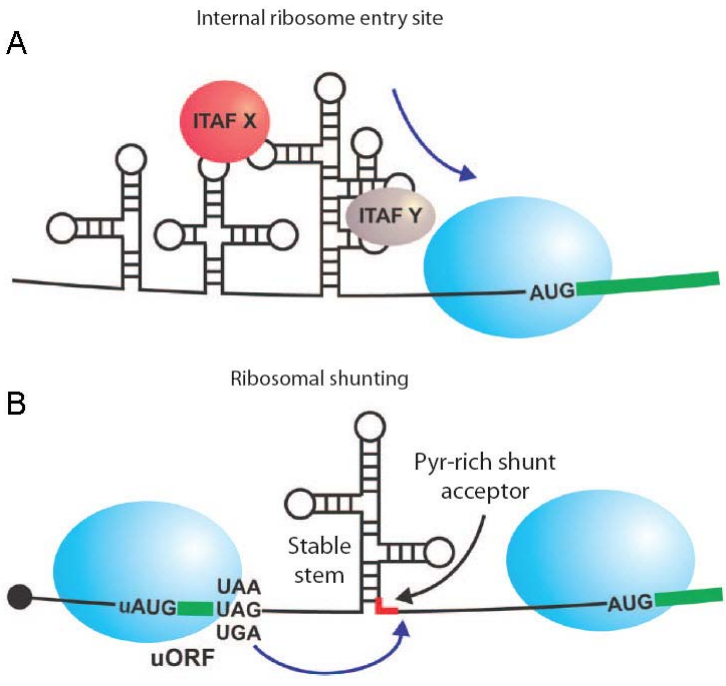


Abb. 1-16: Darstellung alternativer Möglichkeiten der Translations-initiation.
A) 5´Cap-unabhängiger Rekrutierungs-mechanismus des Ribosoms über eine IRES-Struktur („Internal Ribosome Entry Site“). Die Stabilisierung bzw. Rekrutierung kann durch „IRES-specific trans-acting Factors“ ITAFs begünstigt werden.
B) Der Ribosomen-Shunt bezeichnet einen 5´Cap-abhängigen Prozeß der Translations-initiation. Dabei werden weite Teile der 5´UTR, die in Form von Sekundärstrukturen vorliegt, umgangen. Das Ribosom beginnt mit dem Scanning am 5´Cap der 5´UTR und wird durch Sekundärstrukturen in die Nähe des Haupt-ATGs dirigiert. Als Ausgangspunkt kann z.B. ein uORF dienen, als Akzeptor werden Pyrimidin-reiche Sequenzen diskutiert. (Verändert nach Morley & Coldwell, 2008).

Die vorgestellten Möglichkeiten der post-transkriptionellen Expressionskontrolle können sowohl isoliert als auch kombiniert in einer 5´UTR realisiert werden. Durch die hohe Anzahl an Kombinationsmöglichkeiten kann, insbesondere wenn noch zusätzlich alternative 5´- und 3´UTR-Sequenzen, mit unterschiedlichen Eigenschaften, sowie unterschiedliche Längen des Poly(A)-Schwanzes in Betracht gezogen werden (Abb. 1-17), ein hohes Maß an differentieller Translationseffizienz und Gewebespezifität erreicht werden (Wilkie et al., 2003; de Moor et al., 2005; Hughes, 2006; Davuluri et al., 2008; Filipowicz et al., 2008; Smith, 2008; Groppo & Richter).

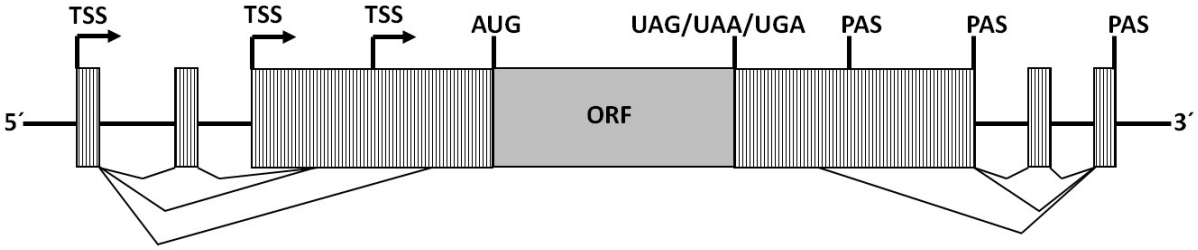


Abb. 1-17: Schematische Darstellung alternativer 5´- und 3´UTR-Sequenzen.
Alternative 5´UTR-Sequenz werden durch Transkriptionsstart an unterschiedlichen TSSs erzeugt. Die alternativen TSSs können innerhalb eines Exons lokalisiert sein, sie können aber auch neue 5´UTR-Exone hinzufügen, die in unterschiedlichen Kombinationen vor ein ATG gesplict werden können. Weiterhin können alternative Polyadenylierungsstellen (PAS) und alternative 3´UTR Splice-Ereignisse zu unterschiedlichen 3´UTR-Längen führen und somit die Komplexität der Regulationsmöglichkeiten weiter erhöhen. Eine weitere Regulationsebene bietet der Poly(A)-Schwanz, dessen Länge je nach Differenzierungzustand oder Gewebe zwischen 20 und 500 Adeninresten variieren kann (Kleene, 1989; de Moor et al., 2005). TSS = Transkriptionsstartstelle, AUG = Translationsstartcodon, ORF = Open Reading Frame, UAG/UAA/UGA = Translationsstopcodons. (Verändert nach Hughes, 2006).

Die Nutzung alternativer TSS, 5´UTRs, Splice-Ereignisse, 3´UTRs, sowie unterschiedlicher Poly(A)-Schwanzlängen zur differentiellen Kontrolle der Translationseffizienz ist vor allem in männlichen Keimzellen weit verbreitet (Braun, 1998; Kleene, 2001, 2003; Steger, 2001; Ehrmann & Elliott, 2005; Elliott & Grellscheid, 2006; Davuluri et al., 2008; Jaroszynski et al., 2011; Rauschendorf et al., 2011).

1.9. Ziel der Arbeit

Ausschlaggebend für die Suche nach *cis*-regulativen Elementen in den 5´ Regionen der homologen Gene *DDX3X* und *DDX3Y*, ist die Proteinexpression in unterschiedlichen Phasen der Human Spermatogenese, trotz offensichtlich austauschbarer Proteinfunktion *in-vitro*. Ursache der unterschiedlichen Genexpression der zwei *DDX3* Allele *DDX3X* und *DDX3Y* scheint der schrittweise Verlust der Gonosomen-Rekombination zu sein, der vor ~ 130 Mio. Jahren einsetzte. Um diese differentielle Expressionskontrolle in der männlichen Keimbahn besser verstehen zu können, wurde die molekulare Basis der *DDX3*(*X/Y*) Genkontrolle

vergleichend zwischen Human X und Y Genkopie, sowie im evolutionären Vergleich mit fünf weiteren Säugerspezies (Schimpanse, Rhesusaffe, Weißbüschelaffe, Rind, Maus) studiert.

Die daraus resultierenden Fragestellungen lauten:

- Gibt es in Human X-Y konservierte *cis*-regulative Sequenzmotive?
- Gibt es Sequenzmotive, die nur typisch X, oder typisch Y sind?
- Gibt es evolutionär konservierte *cis*-regulative Sequenzmotive in der X- und der Y-Linie?
- Gibt es Sequenzmotive die nur spezifisch für Human sind?

Methodisch sollen diese Fragestellungen durch Kombinationen von *in-silico* Analysen, wie z.B. vergleichende TFBS-Kartierungen, und gezielten Experimenten, wie der Kartierung der Transkriptionsstartstellen und funktionellen Luziferase-Reportergenassays, beantwortet werden.
Vor dem Hintergrund der wichtigen Funktionen von *DDX3Y* und *DDX3X* für die Human Spermatogenese, ist die Aufklärung der differentiellen Genexpression in der männlichen Keimbahn von Bedeutung, um zukünftig Patienten mit Fertilitätsproblemen qualitativ besser beraten zu können und im Rahmen von ART („assisted reproductive technologies“) wirkungsvollere Therapien anzubieten.

2. Material

2.1. Geräte und Zubehör

Tab. 2-1: Geräte und Zubehör.

ABI PRISM 3100 Genetic Analyzer	Applied Biosystems, Foster City, USA
Autoklav 3870ELV	Tuttnauer, Breda, Niederlande
Brutschrank für Zellkultur Binder	WTB Binder, Tuttlingen
Brutschrank für Bakterien Controller R2	WTB Binder, Tuttlingen
Biophotometer	Eppendorf, Hamburg
Cryotank GT55	Air Liquide Cryotechnik, Düsseldorf
Digestorium Weiler 0179	Weiler, Wüstenrot
Drigalski-Spatel	neoLab, Heidelberg
Eismaschine AF80	Scotsman, Vernon Hills, USA
Elektrophoresekammer Netzteil (500V, 0,5A)	Fischer, Heidelberg
Elektrophoresekammer Easy-cast B1A, B2	cti, Heidelberg
Gefrierschrank -20˚C	Liebherr, Biberach
Gefrierschrank -80˚C NapCOIL UF600	Napco, Winchester, USA
Gefrierschrank -80˚C NapCOIL UF650	Thermo Fisher Scientific, Waltham, USA
Homogenisator Ultra Turrax T8	IKA Labortechnik, Staufen
Luminometer MicroLumat LB96	Berthold, Bad Wildbad
Magnetrührer MR2000	Heidolph, Schwabach
Magnetrührer mit Heizung RCT	IKA Labortechnik, Staufen
Mikroskop für Zellkultur Diavert	Leitz, Wetzlar
Mikrowelle R239	Sharp, Hamburg
Neubauer-Zählkammer (0,1mm Tiefe, 4x4 Quadrate je 0,0025 mm^2)	LO Laboroptik, Bad Homburg
pH-Meter Calimatic 761	Knick, Berlin
Pipetten Gilson 10/20/200/1000µl	Gilson, Middleton, USA
Pipetten BioPette 10/20/200/1000µl	Labnet, Edison, USA
Pipettierhelfer Accu-Jet Pro	Brand, Wertheim
Schüttelinkubator IH50	Noctua, Mössingen
Schüttelplatte LS10	Gerhardt, Bonn
Schüttelplatte DSG 304M/4	Heidolph, Schwabach
Spectrophotometer ND-1000	NanoDrop, Wilmington, USA
Sterilbank für Zellkultur antair ZKB	Mahl
Sterilbank für Mikrobiologie LaminAir Model 1.8	Holten, Dänemark
T-Gradient PCR-Gerät	Biometra, Göttingen
T1 Thermocycler PCR-Gerät	Biometra, Göttingen
Thermoblock QBT2	Grant Instruments, Shepreth, England

Thermoblock TDB-120	Biosan, Riga, Lettland
Thermomixer comfort	Eppendorf, Hamburg
UV-Imager GelDoc 2000	BioRad, München
UV-Transilluminator IL-200-M	Bachhofer, Reutlingen
Vortex-Schüttler Reax 2000	Heidolph, Schwabach
Waage EW 150-3M	Kern, Balingen
Wasseraufbereitung Milli-Q biocel	Millipore, Eschborn
Wasserbad GFL	Thermolab
Wasserbad 3043	Köttermann GmbH, Uetze/Hänigsen
Zentrifuge Napco 2019R	Napco, Winchester, USA
Zentrifuge Microzentrifuge	Sigma, Taufkirchen
Zentrifuge Biofuge pico	Heraeus, Hanau
Zentrifuge Centrifuge 5417R	Eppendorf, Hamburg
Zentrifuge Cryofuge 5000	Heraeus, Hanau
Zentrifuge Megafuge 1.0	Heraeus, Hanau
Tischzentrifuge Z100M	Labnet, Edison, USA

2.2. Verbrauchsmaterialien

Tab. 2-2: Verbrauchsmaterialien.

Atemschutzmaske 1883 FFP3	3M™, Neuss
Cellstar® 24-Loch-Kulturplatten („24-well plates“)	Greiner, Frickenhausen, # 662160
Cellstar® 96-Loch-Kulturplatten („96-well plates“)	Greiner, Frickenhausen, # 655180
Cellstar® Falcon-Röhrchen (15 u. 50 ml)	Greiner, Frickenhausen, # 188271, 227261
Cellstar® Zellkulturflaschen (250ml, 75cm^2)	Greiner, Frickenhausen, # 658175
Einweg-Skalpell No. 10/15/21	Feather, Osaka, Japan
Impfschlinge 1µl	Greiner, Frickenhausen, # 731161
Klebefolie für Sequenzierplatte	neoLab, Heidelberg, 7-2220
Kryoröhrchen	Nunc, Darmstadt, # 363401
Luminometer Mikroplatte 96-Well Format	Berthold, Bad Wildbad, # 23300
Nitril-Handschuhe Nitra-Tex EP	Ansell, Brüssel, Belgien, # 4400042/70062
Parafilm PM996	Pechiney, Chicago, USA
Pasteur Pipetten 150/230mm	WU-Mainz, Mainz
PCR-Platte für Sequenzierung 96-Well, 0,2ml	Nerbe Plus, Winsen/Luhe, # 04-083-0540
PCR-Reaktionstubes (single cap 8-string 0,2ml)	Biozym, , # 710986
Petrischalen 94x16 mm	Greiner, Frickenhausen, # 633161
Pipettenspitzen 10/200/1000µl	Steinbrenner, Wiesenbach, # 950-001, 950-205, 950-1001
Pipettenspitzen Filter, 10/20/200/1000µl	Stein Labortechnik, Remchingen
PS-Röhrchen für Flüssigkultur (14 ml)	Greiner, Frickenhausen, # 191180

Reaktionsgefäße 0,5/1,5/2ml	Sarstedt, Newton, USA, #
RNase Exitus Plus	AppliChem, Darmstadt, # A7153,1000
Stripette® 5/10/25ml	Corning, Amsterdam, Niederlande, # 4487, 4488, 4489
Verschlußmatte, silikon, 96-Well	Nerbe Plus, Winsen/Luhe, # 04-093-3100
UV-Küvetten	neolab, Heidelberg, # 1310056
Zellkulturflaschen (50ml, 25 cm^2)	Sarstedt, Newton, USA, # 83.1810.002

2.3. Chemikalien

Tab. 2-3: Chemikalien.

2-Propanol (99,5%)	Roth, Karlsruhe, # 9866
Aceton (99,5%)	Roth, Karlsruhe, # 5025
Agar	Roth, Karlsruhe, # 5210.3
Ampicillin Sodium (100mg/ml)	AGS, Heidelberg
Aqua ad iniectabilia (steriles H_2O für PCR)	Braun, Melsungen, # 2351744
Bacto™ Hefeextrakt	Becton Dickinson, Heidelberg, # 212750
Bacto™ Trypton	Becton Dickinson, Heidelberg, # 211705
β-Mercaptoethanol	ICN Biomedicals, Irvine, USA, # 806443
DEPC 96% (Diethylpyrocarbonat)	Serva, Heidelberg, # 18835
DMSO 99,5% (Dimethylsulfoxid)	Roth, Karlsruhe, # A994.2
EDTA (Ethylendiamin-Tetraessigsäure Dinatriumsalz Dihydrat)	Roth, Karlsruhe, # 8043
Essigsäure (100%)	Roth, Karlsruhe, # 3738
Ethanol (99,8%)	Roth, Karlsruhe, # 9065
Ethidiumbromid	Serva, Heidelberg, # 21238
Formaldehyd (37%)	J.T. Baker (jetzt Avantor), Griesheim, # 7040
Formamid (99,5%)	Sigma, Taufkirchen, # 47671
GC-Melt Puffer (5 - 10% einsetzen)	Clontech, Mountain View, USA
Glycerin (100%)	J.T. Baker (jetzt Avantor), Griesheim, # 7044
Kaliumhydrogencarbonat	Ferak, Berlin, # 00898
LE-Agarose	Biozym, Hess. Oldendorf, # 840004
Methanol (99,9%)	Roth, Karlsruhe, # 4627
MOPS (3-Morpholinopropansulfonsäure)	AppliChem, Darmstadt, # A1076
Natriumacetat, wasserfrei	Roth, Karlsruhe, # 6773
Natriumchlorid	Roth, Karlsruhe, # 3957.1
Natriumhydroxid in Plätzchen	Roth, Karlsruhe, # 6771
N,N-Dimethylformamid 99% (DMF)	Sigma, Taufkirchen, # D4551
Orange G	Merck, Darmstadt, # A1404
Sodiumdodecylsulfat (SDS)	Serva, Heidelberg, # 20763
Trition-X 100	Merck, , # 8603

Tris (99,9%)	Roth, Karlsruhe, # 4855
Wasserstoffperoxid H2O2 37%	J.T. Baker (jetzt Avantor), Griesheim, # 7047
X-Gal (5-Bromo-4-Chloro-3-Indoyl-β-D-Galactoside)	Promega, Mannheim, # V394
Xylencyanol	Fluka Chemicals, Buchs, Schweiz, # 95600
Xylol	Roth, Karlsruhe, # 9713

2.4. Fertigsysteme

Alle Fertigsysteme wurden nach Herstellerangaben verwendet.

Tab. 2-4: Fertigsysteme.

ABI BigDye™ Terminator v1.1 Cycle Sequencing RR100	Applied Biosystems, Foster City, USA, # 4336768
ABI High-Di™ Formamide	Applied Biosystems, Foster City, USA, # 4311320
Dual-Luciferase® Reporter Assay	Promega, Mannheim, # E1980
ONE-Glo™ Luciferase Assay System	Promega, Mannheim, # E6110
pGEM®-T Easy Vector System	Promega, Mannheim, # A1360
QIAprep Spin Miniprep Kit	Qiagen, Hilden, # 27106
QIAquick Gel Extraction Kit	Qiagen, Hilden, # 28706
QIAshredder	Qiagen, Hilden, # 79656
RNeasy Minikit	Qiagen, Hilden, # 74106
TnT® T7 Quick Coupled Transcription/Translation Systems	Promega, Mannheim, # L1170

2.5. Enzyme

Enzyme wurden nach Herstellerangaben eingesetzt. Zu allen Enzymen wurden die entsprechenden Puffer genutzt

Tab. 2-5: Enzyme.

Antarctic Phosphatase (AnP)	NEB, # M0289S
GO*Taq* Flexi DNA Polymerase	Promega, # M830B
M-MLV Reverse Transkriptase, H(-), Point Mutant	Promega, # M368C
Pfu DNA Polymerase	Promega, # M774
Proteinase K	
RNase-free DNase Set	Qiagen, Hilden, # 79254
RNase A	Invitrogen, # 12091021
RNase H	Promega, # M428C
Taq DNA Polymerase, rekombinant (5U/µl)	Invitrogen, # 10342-020
T4 DNA Ligase	NEB, # M0202L
T4 DNA Polymerase	Fermentas, # EP0061

Restriktionsendonukleasen:	
BanI	NEB, # R0118S
BbvCI	NEB, # R0601S
BclI	NEB, # R0160S
BglII	NEB, # R0144L
Diluent A	NEB, # B8001S
DpnI	NEB, # R0176S
EcoRI	Fermentas, # ER0271
KpnI	NEB, # R0142L
HindIII	NEB, # R0104L
NheI	NEB, # R0131S
PciI	SibEnzyme über NEB, # V0275S
PmlI	NEB, # R0532S
PvuII	Promega, # R6331
SacI	NEB, # R0156S
XbaI	NEB, # R0145S

2.6. Puffer und Lösungen

Falls nicht anders angegeben, wurden alle Puffer und Lösungen mit gereinigtem Milli-Q Reinstwasser (ddH_2O) angesetzt und ggf. autoklaviert oder sterilfiltriert.

2.6.1. Isolation von genomischer DNA und Total-RNA

Lysis-Puffer (1L):
155mM NH_4Cl (8,29g)
10mM $KHCO_3$ (1g)
0,1mM Na_2EDTA (200µl 0,5M Lösung)
auf 1L mit ddH_2O auffüllen
pH-Wert 7,4 mit 1M HCl einstellen
sterilfiltrieren.

SE-Puffer (1L):
75mM NaCl (4,39g)
25mM Na_2EDTA (8,41g)
auf 1L mit ddH_2O auffüllen
pH-Wert 8,0 einstellen
autoklavieren.

TE-Puffer (1L):
10mM Tris (1,21g)
1mM Na_2EDTA (0,34g)
auf 1L mit ddH_2O auffüllen
pH-Wert 8,0 mit konz. HCl einstellen
autoklavieren.

NaCl gesättigt:
NaCl in dH_2O lösen bis Bodensatz entsteht
autoklavieren.

3 M Na-Acetat (100ml):
24,6g Na-Acetat (anhydrid) in 80ml ddH_2O anlösen
pH-Wert auf 7,0 mit Eisessig einstellen
auf 100ml mit ddH_2O auffüllen.

Proteinase K:
Proteinase K (10mg) in 1ml SE-Puffer lösen. Lagerung -20°C.

20 % SDS (100 ml):
20g SDS (Natrium Dodecyl Sulfat) in 100ml ddH_2O lösen.

RLT + ß-Mercaptoethanol:
10µl ß–Mercaptoethanol pro 1ml RLT (RNA lysis thiocyanate)-Lysis-Puffer aus Qiagen “RNeasy Mini Kit”.

70 % Ethanol:
Ethanol abs. mit ddH_2O (70 zu 30) verdünnen.

80 % Ethanol:
Ethanol abs. mit ddH_2O (80 zu 20) verdünnen.

2.6.2. RNA-Gel

DEPC-H_2O (1L):
1ml DEPC (1 ‰) pro 1L ddH_2O
unter Abzug 6 h rühren
autoklavieren.

50 mM Na-Acetat (1L):
4,1g in 1 L DEPC-H_2O (1 ‰ DEPC) lösen
autoklavieren.

500 mM EDTA (1L):
18,612g EDTA in 1L DEPC-H_2O
pH-Wert 8.0 mit 10M NaOH einstellen
autoklavieren.

5x MOPS-Puffer (1L):
20,6g MOPS in 800ml 50mM Na-Acetat lösen
pH-Wert 7.0 mit frischer 2M NaOH einstellen
10ml 500 mM EDTA zugeben
auf 1L mit DEPC-H_2O auffüllen
sterilfiltrieren und lichtgeschützt lagern.

RNA-Gel-Ladepuffer:
100µl Formamid (100%)
40µl Formaldehyd (37%)
20µl 5x MOPS
1µl Ethidiumbromid (10 mg/ml)

Bromphenolblaulösung:
50% Glycerol in H_2O
1mM EDTA pH-Wert 8.0
0,25% Bromphenolblau
0,25% Xylencyanol

Vorbereitung der Gelkammer:
3% H_2O_2 zur Sterilisation der Gelkammer.

2.6.3. Gelelektrophorese

50x TAE:
242g Tris (2M)
18,6g Na_2EDTA (50mM)
57,1 ml Eisessig
auf 1L mit ddH_2O auffüllen.

Ladepuffer Orange G:
20ml Glycerin (87%)
1ml 1M Tris (pH 7,5)
0,1g Orange G
auf 100ml mit ddH_2O auffüllen.

1x TAE:
200ml 50x TAE in 10L ddH_2O.

2.6.4. Kulturmedien und -Lösungen zur Bakterienkultivierung

LB (Luria Bertani)-Flüssigmedium (1L):

NaCl	10g
Bacto Trypton	10g
Hefeextrakt	5g

auf 1L mit ddH_2O auffüllen, pH 7,0 einstellen und autoklavieren. Lagerung bei 4°C.

LB-Agarplatten:

In vorbereitetes LB-Flüssigmedium zusätzlich 20g Agar/L Medium einwiegen und autoklavieren. Nach dem Autoklavieren auf ca. 50 - 55°C abkühlen lassen und vor dem Gießen der Platten 500µl Ampicillin-Lösung (100mg/ml) pro Liter LB-Agar zugegeben. Zur Blau-Weiß Selektion transformierter Bakterienkolonien wurden bei Bedarf vor dem Gießen der Platten 800µl X-Gal (50 mg/ml) pro Liter LB-Agar zugegeben.

X-Gal (50 mg/ml):
In 100% Ethanol lösen und sterilfiltrieren
Lagerung bei -20°C.

Ampicillin (100 mg/ml):
In ddH_2O lösen und sterilfiltrieren
Lagerung bei -20°C.

2.7. Gewebe

***Homo sapiens*:**

Hodengewebe (Hodenbiopsat) mit normaler Spermatogenese (Laborcode: Testis b) stand in der Arbeitsgruppe von Herrn Prof. Dr. rer. nat. P.H. Vogt, Universitätsfrauenklinik Heidelberg, zur Verfügung. DNA und RNA aus männlichen Leukozyten wurden aus Frischblut RAUMAR aufbereitet. DNA und RNA aus weiblichen Leukozyten wurden aus Frischblut BENULI aufbereitet.

***Pan troglodytes*:**

Schimpansengewebe wurden über das Projekt EUPRIM-Net (European Primate Network; EU contract RII3-026155 of the 6th Framework Programme) vom Biomedical Primate Research Centre BPRC (Rijswijk, Niederlande; namentlich Dr. Gaby Doxiadis und Dr. Ivanela Kondova) erhalten. Der Transfer von Schimpansengewebe ist durch CITES-Antrag (Convention on International Trade in Endangered Species of Wild Fauna and Flora), beim zuständigen niederländischen Ministerium für Landwirtschaft, Natur und Lebensmittelqualität, genehmigt.

- 12 Jahre altes männliches fertiles Tier Japie (*04.03.1988, †29.06.2000) mit Mikrochiptransponder Nr. 0014335A8: Hoden, Niere, Leber, Milz.

***Macaca mulatta*:**

Rhesusaffengewebe wurden über das Projekt EUPRIM-Net (European Primate Network; EU contract RII3-026155 of the 6th Framework Programme) vom Biomedical Primate Research Centre BPRC (Rijswijk, Niederlande; namentlich Dr. Gaby Doxiadis und Dr. Ivanela Kondova) erhalten.

- 12 Jahre altes männliches fertiles Tier : Hoden, Niere, Leber, Milz.

***Callithrix jacchus*:**

Weißbüschelaffengewebe wurden über das Deutsche Primatenzentrum (Göttingen; namentlich Dr. Penny Nayudu) bezogen und standen in der Arbeitsgruppe von Herrn Prof. Dr. rer. nat. P.H. Vogt, Universitäts-frauenklinik Heidelberg, zur Verfügung.

- männliches Tier Paco (# 9652): Hoden, Niere, Leber.
- Total-RNA aus männlichen Leukozyten stand in der Arbeitsgruppe von Herrn Prof. Dr. rer. nat. P.H. Vogt, Universitätsfrauenklinik Heidelberg, zur Verfügung.

***Bos taurus*:**

Rinderfrischgewebe (Hoden, Niere, Leber, Milz, Blut) wurden über das FVZ Mannheim (Fleischversorgungszentrum, Schlachthofstraße 21, 68165 Mannheim) bezogen.

***Mus musculus*:**

Maus Hodengewebe und Total RNA aus männlichen Leukozyten standen in der Arbeitsgruppe von Herrn Prof. Dr. rer. nat. P.H. Vogt, Universitätsfrauenklinik Heidelberg, zur Verfügung.

2.8. Zellkulturen

2.8.1. Medien und Zusätze für Zellkulturen

Tab. 2-6: Medien und Zusätze für Zellkulturen (PAA* = PAA Laboratories GmbH, Pasching, Österreich).

Antibiotic/Antimyotic Solution (Penicillin, Streptomycin, Amphotericin B)	PAA*, # P11-002
β-Mercaptoethanol	Gibco-BRL, # 31350
DMEM (low Glucose 1g/L) mit L-Glutamin	PAA, # E15-806
DMEM (low Glucose 1g/L) ohne L-Glutamin	PAA, # E15-005
DMSO (99,9%)	Sigma, Taufkirchen, # D8418
Dulbecco´s PBS ohne Ca u. Mg	PAA, # H15-002
Fötales Kälberserum (FBS)	Gibco/Invitrogen, Karlsruhe, # PL16141079
Metafectene Transfektionsreagenz für Säugerzellen	Biontex, Martinsried/Planegg, # T020
RPMI 1640 mit stabilem Glutamin	PAA, # E15-885
Trypanblau zur Zellzahlbestimmung	Sigma, Taufkirchen, # T8154
10x Trypsin-EDTA Lösung	PAA, # L11-003

2.8.2. Human Zellinie

Die EP2102 Zellinie ist eine Human testikulär, embryonale Karzinomazellinie (Andrews, 1984; Sheffield UK). EP2102 bildet, bei 37°C, 5% CO_2 und 80% Luftfeuchtigkeit, einen Monolayer in DMEM (low glucose) + 10% FBS + 0,1mM (= 0,1%) β-Mercaptoethanol + 1% Antibiotic/Antimyotic Mix. Zur Verfügung gestellt durch Dr. Ewa Rajpert-De Meyts, Kopenhagen Rigshospitalet, Dänemark.

2.8.2. Primaten Zellinien

***Pan troglodytes*:**

- männliches fertiles Tier (Dirk1): immortalisierte B-Zellen.
- weibliches fertiles Tier (Natascha): immortalisierte B-Zellen.

Schimpansen B-Zellinien wurden im BPRC mit EBV (Epstein-Barr-Virus = Humanes-Herpes-Virus 4) transformiert. Die transformierten B-Zellen wachsen in Haufen und benötigen hohe Zelldichten. Sie wachsen in RPMI 1640 (mit stabilem Glutamin) + 10% FBS + 1% Antibiotic/Antimyotic Mix.

***Macaca mulatta*:**

- männliches fertiles Tier (# 97058): immortalisierte B-Zellen.
- weibliches fertiles Tier (# 99031): immortalisierte B-Zellen.

Macaca B-Zellinien wurden im BPRC mit HVP (Herpes Virus Papio) transformiert. Die transformierten B-Zellen wachsen in Haufen und benötigen hohe Zelldichten. Sie wachsen in RPMI 1640 (mit stabilem Glutamin) + 10% FBS + 1% Antibiotic/Antimyotic Mix.

2.9. Chemo-kompetente Bakterienzellen

***E. coli* DH5α** (Invitrogen, Karlsruhe # 18265-017).
Genotyp: F^-, φ80*lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *rec*A1, *end*A1, *hsd*R17(rk^-, mk^+), *pho*A, *sup*E44, *thi*-1, *gyr*A96, *rel*A1 λ^-.

***E. coli* JM109** (Promega, Mannheim # L2001).
Genotyp: *end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (rk–, mk+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *laq*IqZΔM15].

Beide verwendeten *E. coli* Stämme sind defizient für die unspezifische Endonuklease EndA (*end*A^-). Dies ist notwendig für eine hohe Plasmidstabilität. Beide verwendeten *E. coli* Stämme sind positiv für die dam Methylase (dam^+). Dies ist notwendig für die „QuickChange site-directed Mutagenesis“. Beide Stämme können für eine Blau-Weiß Selektion ohne IPTG genutzt werden.

2.10. Nukleotide und Nukleinsäuren

Die Human BAC-Klone **RP13-13A3** (*DDX3X*, GenBank Acc.No. AL391647.16) und **475I1** (*DDX3Y*, GenBank Acc.No. AC004474.1) standen in der Arbeitsgruppe von Herrn Prof. Dr. rer. nat. P.H. Vogt, Universitätsfrauenklinik Heidelberg, zur Verfügung.

Tab. 2-7: Nukleotide.

Oligo$(dT)_{15}$	Bioron, Ludwigshafen, # 509001
dNTPs	Bioron, Ludwigshafen, # 110031

2.11. DNA-Größenstandards

Tab. 2-8: DNA-Größenstandards.

GeneRuler™ 50bp DNA Ladder	Fermentas, St. Leon-Rot, # SM0371
GeneRuler™ 100bp DNA Ladder Plus	Fermentas, St. Leon-Rot, # SM0322
GeneRuler™ 1kb DNA Ladder	Fermentas, St. Leon-Rot, # SM0311
Lambda DNA/*HindIII* Marker, 2	Fermentas, St. Leon-Rot, # SM0101
6x MassRuler™ Loading Dye Solution	Fermentas, St. Leon-Rot, # R0629

2.12. Vektoren

Tab. 2-9: Verwendete kommerzielle Vektoren und modifizierte Vektoren.

Vektor Bezeichnung	**Vektorlänge in bp**	**Reportergen**	**Promotor**	**Bezugsquelle**
pGEM-T easy	3015	*lacZ* (β-Galaktosidase)	T-7 RNA Polymerase-Promoter	Promega, Mannheim
pUC18	2686	*lacZ* (β-Galaktosidase)	---	Invitrogen, Karlsruhe
pGL4.10 (*luc*2)	4242	Firefly Luciferase *Luc*2 (Codon-Usage optimiert für Säuger Zellkultur)	Endogen ohne Promoter	Promega, Mannheim
pGL4.13 (*luc*2/SV40)	4641	Firefly Luciferase *Luc2*	Simian Virus 40 (SV40)	Promega, Mannheim
pGL4.73 (hRluc/SV40)	3921	*Renilla* Luciferase hRluc (Codon-Usage optimiert für Säuger Zellkultur)	Simian Virus 40 (SV40)	Promega, Mannheim
pGL4.74 (hRluc/TK)	4237	*Renilla* Luciferase hRluc	„Herpes simplex Virus“Thymidin Kinase (HSV-TK)	Promega, Mannheim
pGL4.75 (hRluc/CMV)	4281	*Renilla* Luciferase hRluc	Cytomegalievirus (CMV)	Promega, Mannheim
Luciferase T7 Control DNA	4331	Firefly Luciferase *Luc*+ (pGL2 Generation)	T-7 RNA Polymerase-Promoter	Promega, Mannheim
pUC18	2686	Modifiziert mit Firefly Luciferase *Luc*2	Modifiziert mit T-7 RNA Polymerase-Promoter	Invitrogen, Karlsruhe; Modifikation in dieser Arbeit

2.13. Oligonukleotide für PCR und RT-PCR

Die Primer wurden von der Firma Thermo Electron GmbH (Ulm) bezogen. Die Tabellen mit den verwendeten Primersequenzen zur Kartierung der *DDX3X* und *DDX3Y* TSS befinden sich im Anhang (Tab. A-38 bis A-40). Die Tabelle mit den verwendeten Primersequenzen zur Insertkontrolle und Sequenzierung befindet sich im Anhang (Tab. A-41). Die Tabelle mit den verwendeten Primersequenzen zur Generierung der pGL4 Luziferase-Reportergenvektoren befindet sich im Anhang (Tab. A-42). Die Tabelle mit den verwendeten Primersequenzen zur Generierung der pUC18 Vektoren für die *DDX3X in-vitro* Translation befindet sich im Anhang (Tab. A-43).

2.14. Software und Datenbanken

Für die Datenverarbeitung wurden die gängigen Microsoft Office-Programme (Suite 2010) verwendet. Als Grafikprogramm wurde Adobe Photoshop Elements (Vers. 1) verwendet. Für alle Arbeiten mit Auswertungs-Software (Geldokumentation, Luminometer) wurden die vom Hersteller mitgelieferten Software-Pakete benutzt. Des Weiteren wurde das Sequenzanalyse Programm „Sequencher 4.2“ der Firma Gene Codes eingesetzt. Mit Hilfe der in Tabelle 2-10 dokumentierten Datenbanken wurden die DNA-, RNA-, und Proteinsequenzen untersucht.

2.14.1. Genomatix-Software zur Identifizierung von Transkriptionsfaktor-bindestellen

Die wichtigsten Fakten zum Verständnis der Transkriptionsfaktor-Analyse seien hier in Kürze angeführt. Die meist kurzen und variablen TFBS werden üblicherweise über zwei Ansätze detektiert, da es nicht EINE bestimmte Bindungssequenz gibt. Einerseits werden IUPAC-Konsensussequenzen verwendet, die mehrdeutige Symbole (z.B. R = A oder G) für die Beschreibung der variablen TFBS nutzen. Andererseits werden Gewichtungsmatrizen („Weight Matrix“), auch „Nukleotid-Verteilungsmatrix“ genannt, angewendet. Gewichtungsmatrizen werden durch eine zweidimensionale Tabelle dargestellt, in der für jede Position der TFBS eine Nukleotidpräferenz ausgedrückt wird (Stormo, 2000). Idealerweise sind diese Matrizen anhand von funktionell charakterisierten TFBS für einen gegebenen TF erstellt. Ein aus dieser Matrix abgeleiteter Qualitätswert („Score“) kann genutzt werden, um die relative Bindungseffizienz eines bestimmten TFs an eine DNA-Sequenz abzuschätzen. Die Bindungsaffinität eines TFs und somit die biologische Relevanz eines TFs wird über einem bestimmten Schwellenwert erwartet.

Das Genomatix-Programm „MatInspector“ (Vers. 8.0, 04/2009; Cartharius et al., 2005) nimmt in seinen Analysen eine Gewichtung der möglichen Bindungsfähigkeit von TFs an einen bestimmten Sequenzabschnitt vor, und drückt diese in „Core-“, und „Matrix-Werten“ aus. Der „Optimized-Wert“ (Opt-Wert) stellt einen individuellen unteren Grenzwert für jede TF-Matrix unabhängig von der eingegebenen Sequenz dar und soll dazu beitragen, vor allem falsch-positive TFBSs zu reduzieren. Längere TF-Matrizen haben üblicherweise, trotz hoher Konserviertheit, einen niedrigeren Matrix-Wert, da mehr Mismatches vorkommen können. Der Opt-Wert soll auch verhindern, daß lange TF-Matrizen, die häufig einen festen Matrix-Grenzwert nicht erreichen, durch die Grenzwerteinstellung weggelassen werden. Somit wägt der Opt-Wert zwischen hochspezifischen oder langen TF-Matrizen und unspezifischen oder kurzen TF-Matrizen ab. In der MatInspector-Beschreibung ist angegeben, daß ein Matrix-Wert für eine potentiell funktionelle TFBS oberhalb des dazugehörigen Opt-Wertes liegen sollte. Die MatInspector-Standardeinstellung ist zudem so gewählt, daß nur TFs mit einem TF-Core-Wert von mind. 0,75 angegeben werden. Als TF-Core sind bei MatInspector meist die vier am höchsten konservierten Nukleotide einer TF-Bindesequenz definiert. Werte unterhalb dieser Grenze werden nicht berücksichtigt, da für solche TFs die Bindungswahrscheinlichkeit geringer ist. Von Genomatix wird angegeben, daß Matrix-Werte oberhalb von 0,85 eine gute Übereinstimmung zur Konsensussequenz liefern. Da eine TF-Matrix die gesamte Bindesequenz eines TFs beschreibt, wird dem gesamten Matrix-Wert ein höherer Stellenwert, als nur dem reinen TF-Core-Wert, beigemessen. MatInspector gibt somit drei Parameter aus (TF-Core > 0,75, Matrix-Wert > 0,85 und Matrix-Wert ≥ Opt -Wert), die auf eine funktionelle Bindung hinweisen können und anhand der TF-Analyse dann experimentell zu überprüfen sind.
Aktuell verfügt das Genomatix-Programm MatInspector über eine Sammlung von 5747 Vertebraten-TFs und deren Bindestellen, die in 727 Gewichtungsmatrizen eingeteilt sind, die wiederum in 170 Matrix-Familien, sortiert sind.

2.14.2. Erstellung von Sequenzalignments mit den Programmen mVISTA, ClustalW und DiAlignTF

Das Programm mVISTA (Frazer et al., 2004) nutzt für die Sequenzalignments das Programm LAGAN (Limited Area Global Alignment of Nucleotides), für paarweise Alignments und MLAGAN (Multi-LAGAN) für Alignments multipler Spezies (Brudno et al., 2003). Beide Programme arbeiten mit dem CHAOS Algorithmus (Brudno & Morgenstern 2002), der lange, globale Alignments in drei Schritten berechnet. Zunächst werden viele kurze, lokale Alignments gebildet, diese werden in einem zweiten Schritt in Ketten miteinander verknüpft, wobei die lokalen Alignments als Anker dienen. Dieser Schritt wird solange wiederholt, bis ein grobes Alignment aller Anker erstellt ist. Im letzten Schritt wird das globale Alignment aus den so erhaltenen Ketten erstellt.

Für die Berechnung der Sequenzkonserviertheit nutzt VISTA eine “sliding window” Methode, bei der die Sequenzidentität in einem Fenster von 100 bp berechnet wird und anschließend immer um ein Nukleotid verschoben wird, um die Sequenzhomologie erneut zu berechnen. Das VISTA-Alignment unterscheidet sich etwas von den EMBOSS ClustalW- und Genomatix DiAlign-Ergebnissen ab, da jedes Programm einen eigenen Algorithmus zur Berechnung der Sequenzhomologien verwendet. mVISTA legt besonderen Wert auf die korrekte Darstellung langer und entfernt homologer Sequenzen. Funktioniert aber durch den CHAOS Algorithmus auch akkurat für kürzere Sequenzen (Brudno & Morgenstern 2002). ClustalW ist für kurze und relativ nahverwandte Sequenzen konzipiert. Das Genomatix-Programm „DiAlign professional TF“ (Vers. 3.1.4, 02/2010) legt besonderen Wert auf die Erkennung konservierter TFBS, was ebenfalls ein sehr akkurates Sequenzalignment zweier oder mehrerer Sequenzen (multiple sequence alignment, MSA) benötigt und arbeitet daher sehr stark mit konservierten Sequenzblöcken anstatt mit einzelnen Nukleotiden (Hertz & Stormo, 1999). Die Sequenzen werden in konservierte Sequenzblöcke, sogenannte Diagonalen, zerlegt und einzeln gewichtet. Das Gesamtalignment wird schließlich aus den einzelnen Blöcken mit jeweils dem maximalen Gewichtungswert zusammengesetzt. DiAlign verwendet keine Strafpunkte für Sequenzlücken („gap penalties“), was das Programm geeignet macht, um lokale Sequenzähnlichkeiten zu detektieren (Morgenstern, 1999). In dieser Arbeit wird folglich für die Suche nach cTFBS das DiAlign-Programm verwendet.

2.14.3. CpG-Island Analyse

Die 2,5 kb langen Sequenzen der *DDX3Y* und *DDX3X* Promoterregionen wurden mit zwei unterschiedlichen Programmen, „CpG Island Searcher“ (Vers. 29.10.2004) und „EMBOSS CpGPlot“ zur Vorhersage von CGIs analysiert. Beide Programmen arbeiten nach der Methode des „Sliding-window Approach“ (SWA), die mit definierten Fenstergrößen in definierten Schritten die Sequenzen, nach den Parametern (C+G)-Gehalt, Verhältnis von CpG beobachtet zu CpG erwartet („O/E CpG“) und der Länge eines CGIs, scannen.

Das Programm „CpG Island Searcher“ verwendet die von Takai & Jones (2002) eingeführten, stringenten Kriterien für ein CpG-Island (mind. 500 bp Länge, mind. 55% G+C Gehalt und mind. einen Wert von 0.65 für das Verhältnis von O/E CpG). Diese strenge Definition schließt die meisten (C,G)-reichen *Alu*-Elemente aus, da diese meist ~ 300 bp lang sind. “CpGPlot” (Rice et al., 2000) verwendet die relaxierten Parameter (mind. 200 bp Länge, mind. 50% G+C Gehalt und mind. einen Wert von 0.60 für das Verhältnis von O/E CpG) von Gardiner-Garden & Frommer (1987), welche für die Analyse bereits bekannter Promoterregionen von Interesse sind, da kürzere CpG Abschnitte (CpG-Islets = Inselchen) nicht zu einem langen CGI vereint werden. Solche kurzen CpG-Islets sind häufig unmethyliert und konserviert und können mit alternativen TSSs assoziiert sein (Hackenberg et al., 2006, 2010).

Tab. 2-10: Verwendete Internet-basierte Datenbanken und Programme.

Programm	Programmversion	www.-Adresse	Verwendung	Referenz
BIOBASE (TRANSFAC, Match)	Vers. 12.1	http://www.biobase-international.com/pages/index.php	TFBS-Analyse	Kel et al., 2003; Matys et al., 2006; Wingender, 2008
CpG Island Searcher	29.10.2004	http://cpgislands.usc.edu/	CpG-Island Analyse	Takai & Jones, 2002
DBTSS	6.01	http://dbtss.hgc.jp/	5´Transkriptenden	Wakaguri et al., 2008
EMBL-EBI (EMBOSS-Programme) (ClustalW, CpGPlot)	2006 - 2011	http://www.ebi.ac.uk/ http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html	Sequenzalignment und CpG-Island Analyse	Chenna et al., 2003; Gardiner-Garden & Frommer, 1987
ENSEMBL	2006 - 2011	http://www.ensembl.org/index.html	Genomische und RNA-Sequenzen, BLAST	Flicek et al., 2011
ExPASy Proteomics Server	2006 - 2011	http://www.expasy.org/	Proteinsequenzen und Analyse	Gasteiger et al., 2003
Genomatix (MatInspector, ElDorado, DiAlignTF, ModelInspector)	MatInspector Vers. 8.0.4, ElDorado Vers. 4.9	http://www.genomatix.de/en/index.html	TFBS-, S/MAR- und Genlocus-Analyse. Sequenzalignment	Cartharius et al., 2005
Human Protein Atlas	6.0	http://www.proteinatlas.org/index.php	Proteinexpression	Uhlén et al., 2005; Brtglung et al., 2008
MARFinder	2006 - 2011	http://genomecluster.secs.oakland.edu/marwiz/	S/MAR-Analyse	Singh et al., 1997
mfold	Vers. 3.2	http://mfold.rna.albany.edu/	RNA-Sekundärstrukturen	Zuker, 2003
NCBI (PubMed, Nucleotide, BLAST, Gnomon, Splign, dbSNP)	2006 - 2011	http://www.ncbi.nlm.nih.gov/guide/	Literaturrecherche, genomische und RNA-Sequenzen, BLAST	Sayers, et al., 2011
NEB (NEB Cutter)	V2.0	http://tools.neb.com/NEBcutter2/index.php	Restriktionsenzymschnittstellen	Vincze et al., 2003
PrimerX	2006 - 2011	http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi	Primerdesign für „site-directed mutagenesis”	Lapid & Gao
QGRS-Mapper	2006	http://bioinformatics.ramapo.edu/QGRS/index.php	G-Quadruplex Strukturen	Kikin et al., 2006
RepeatMasker	open-3.3.0	http://repeatmasker.org	Identifizierung repetitiver DNA	Smit, Hubley, Green
Sequence Manipulation Suite	2006 - 2011	http://www.bioinformatics.org/sms2/index.html	ORF-Analyse, Nukleotidverhältnisse; Nukleotid Ambiguity Code	Stothard, 2000; Cavener, 1987
TRACTS	2006 - 2011	http://bip.weizmann.ac.il/tracts/tracts.html	Analyse homomerer und binärer Sequenzen	Gal et al., 2003
UniProt	2010_09	http://www.uniprot.org/	Protein-Annotation	Jain et al., 2009
VISTA	2006 - 2011	http://genome.lbl.gov/vista/index.shtml	Vergleichende Genomik, konservierte Sequenzen	Visel et al., 2007; Prabhakar et al., 2006
WebSat	2006 - 2011	http://wsmartins.net/websat/	Mikrosatelliten	Martins et al., 2009

3. Methoden

3.1. Molekularbiologische & Mikrobiologische Methoden

Die folgenden Methoden sind Standardmethoden in der Molekularbiologie und wurden anhand von Protokollen aus der Methodensammlung von (Sambrook & Russel, 2001) durchgeführt. Die Verwendung von Fertigsystemen erfolgte, soweit nicht anders erläutert, nach Anleitung ihrer Hersteller. Falls Modifikationen zu den Standardprotokollen durchgeführt wurden, sind diese angegeben.

3.1.1. Isolierung genomischer DNA und Total-RNA

Das hier beschriebene Zellyse-Protokoll wurde sowohl für DNA, als auch RNA verwendet. Beide Moleküle wurden aus EDTA-Vollblut, PBS-Zellpellets und Kryogewebe isoliert. Das Protokoll ist prinzipiell in mehrere Schritte unterteilt: 1) Lyse und Entfernung der kernlosen Erythrozyten, 2) Entfernung der Zell- und Kernmembran, sowie von Lipiden und Proteinen, 3) Präzipitation von DNA und RNA und 4) Resuspendierung der reinen Nukleinsäuren in einem geeigneten Puffer.

3.1.1.1. Isolierung aus Vollblut

Lyse:

- 10ml Vollblut werden mit 30ml eiskaltem (4˚C) Lysispuffer in einem 50ml Falcon-Röhrchen gemischt.
- Zur Lyse der Erythrozyten Falcon-Röhrchen 30 min auf Eis inkubieren.
- 10 min bei 4˚C zentrifugieren (1.200 rpm) zur Sedimentation der Lymphozyten. Überstand abgießen.
- Lymphozytenpelett in 10ml eiskaltem Lysispuffer resuspendieren. 5 min bei 4˚C zentrifugieren (1.200 rpm). Vorgang wird so lange wiederholt, bis Pellet weiß und frei von Erythrozyten ist.

Diese Schritte sind identisch für die Isolation von DNA, sowie RNA.

Resuspension für DNA Präparation:

- Resuspension des Lymphozytenpeletts in 5 ml eiskaltem SE-Puffer.
- 5 min bei 4˚C zentrifugieren (1.200 rpm). Überstand abgießen.
- Resuspension des Lymphozytenpeletts in 5 ml SE-Puffer (Raumtemperatur). 40µl Proteinase K und 250µl 20% SDS zugeben.
- Vorsichtig mischen bis Schlieren entstehen.
- Inkubation bei 37˚C (20 - 24 h).
- Bis zur Weiterverarbeitung bei 4˚C lagern.

Resuspension für RNA Präparation:

- Resuspension des Lymphozytenpeletts in 5 ml eiskaltem Lysispuffer.
- 5 min bei 4˚C zentrifugieren (1.200 rpm). Überstand abgießen.
- Resuspension des Lymphozytenpeletts in RLT (~ 200µl/ml Vollblut) + β-Mercaptoethanol (10µl/ml RLT).
- 1 min im Vortex-Schüttler mischen.
- Aliquots mit je 650µl herstellen und bei -80˚C lagern.

DNA-Isolierung per Salzfällung:

- 10ml warmen SE-Puffer zugeben. 3,3ml gesättigte NaCl-Lösung (steril) zugeben und 15 sec. im Vortex-Schüttler mischen.
- 15 min bei Raumtemperatur zentrifugieren (3.000 rpm). Überstand in ein neues Falcon-Röhrchen überführen. Die Zentrifugation wird so lange wiederholt, bis der Überstand klar ist. Nach jeder Zentrifugation wird der Überstand in ein neues Falcon-Röhrchen überführt.
- Zur Fällung der DNA werden 0,1 Volumenanteile 3 M Na-Acetat (pH 7,0) und ein Volumenanteil Isopropanol zugegeben. Vorsichtig mischen bis ein DNA-Faden sichtbar wird.
- DNA-Faden aufnehmen und in 500µl 80% Ethanol gewaschen. DNA-Faden in einem Reaktionsgefäß trocknen lassen, so daß kein Ethanol mehr im Gefäß verbleibt.
- Je nach Größe des Fadens wird dieser in 300 - 1000µl TE-Puffer 10/1 aufgenommen und über Nacht bei 4˚C auf einem Rotor gelöst.
- Zur Konzentrationsmessung der DNA wird diese für 10 min auf 65˚C erhitzt und auf Eis wieder abgekühlt. Eine 1:100 Verdünnung der DNA wird im Photometer gemessen.

RNA-Isolierung nach Qiagen „RNeasy Minikit“ Protokoll:

- Die RNA in RLT-Puffer + β-Mercaptoethanol schnell bei RT auftauen und kurz im Vortex-Schüttler mischen. Probe für 10 min bei 37˚C inkubieren.
- Die nachfolgende Isolierung erfolgte nach dem Qiagen „RNeasy Minikit“ Protokoll.
- Zur Konzentrationsmessung der RNA wird 1µl der Probe mit dem NanoDrop-1000 Photometer bestimmt.

3.1.1.2. Isolierung aus Kryogewebe

Die verwendeten Gewebe wurden bei -80˚C gelagert. Mit sterilen Geräten wurden Gewebeproben in ein Reaktionsgefäß mit 500µl eiskaltem Lysispufferüberführt. Das Gewebe wird mit Hilfe eines Gewebe-Homogenisators homogenisiert. Die nachfolgenden Schritte entsprechen den oben beschriebenen Protokollen für DNA- und RNA-Isolierung.

3.1.1.3. Isolierung aus PBS-Pellets

Die Isolierung von DNA und RNA aus PBS-Zellpellets entspricht den Protokollen für DNA- und RNA-Isolierung aus Vollblut. Die Lyse entfällt.

Das tiefgefrorene PBS-Zellpellet wird direkt in SE-Puffer (DNA), bzw. Lysispuffer (RNA) resuspendiert.

3.1.1.4. Kontrolle der isolierten Total-RNA

Zum Ausschluß einer RNA-Degradation wird die isolierte Total-RNA per Formaldehyd Agarosegel-Elektrophorese kontrolliert. Eine intakte RNA zeigt in einem Ethidiumbromid-gefärbten Agarosegel mindestens zwei deutlich getrennte Banden. Die größere Bande entspricht der 28S-rRNA, die untere der 18S-rRNA. Die 28S-rRNA sollte bei guter RNA-Qualität doppelt so stark sein, wie die 18S-rRNA (2:1 Verhältnis). Ein Bandenschmier ist ein deutliches Anzeichen für bereits erfolgten Abbau der RNA. Die 5S-rRNA Bande ist meist nicht zu sehen.

Agarosegel (1,2%):

- 0,36g LE-Agarose in 24ml ddH_2O aufkochen und bis auf 50 - 60˚C abkühlen lassen.
- 6ml 5x MOPS dazugeben und mischen.
- Gel in RNase-freier Gelkammer (H_2O_2 Behandlung) gießen.
- Polymerisation des Gels für mind. 15 min.
- Nach Polymerisation Gelkammer mit 1x MOPS Laufpuffer (300ml) befüllen.

RNA:

- 0,5µg RNA mit autoklaviertem ddH_20 auf 10µl auffüllen.
- 10µl RNA-Gelladepuffer zugeben und mischen.
- Probe bei 65˚C inkubieren und dann auf Eiswasser stellen. Zentrifugieren.

Gellauf:

- Vorbereitete RNA-Probe auftragen.
- Freie Geltasche mit 5µl Bromphenolblau zur Kontrolle der Lauflänge beladen.
- Gel bei 60 - 65 V laufen lassen und mit UV-Gelimager dokumentieren.

3.1.2. cDNA-Synthese

Die cDNA (complementary DNA) wurde mit Hilfe des Promega M-MLV Reverse Transkriptase-Systems hergestellt. Eine intakte RNA ist Voraussetzung für die Synthese von 5´ vollständiger cDNA Moleküle. Die Erststrangsynthese wird von der M-MLV RT katalysiert. Für die cDNA-Synthese wird ein Oligo$(dT)_{15}$-Primer verwendet, der an den Poly(A)-Schwanz der eukaryotischen mRNAs bindet. Da nur reife mRNAs über einen Poly(A)-Schwanz verfügen, wird mit diesem Schritt auf diese RNA-Spezies selektioniert.

Das finale Gesamtvolumen der cDNA-Reaktion beträgt 26µl.

1) RNA/Oligo(dT)$_{15}$-Primer-Mix:

- X µl Total-RNA (1000 ng - 1300 ng)
- 2 µl Oligo(dT)$_{15}$-Primer
- X µl steriles H_2O, auf 14µl auffüllen
- Mix für 5 min bei 70°C inkubieren
- Mix für 5 min auf Eis abkühlen lassen.

2) Reaktionsmix:

- 5 µl RT-Puffer
- 1,25 µl dNTP-Mix
- 3,75 µl steriles H_2O
- 1 µl M-MLV RT
- 14 µl RNA/Oligo(dt)$_{15}$-Primer-Mix zugeben
- Für 10 min bei 40°C inkubieren
- Für 50 min bei 42°C inkubieren
- Reaktion für 15 min bei 75°C beenden
- 1 µl RNase H zugeben
- Für weitere 20 min bei 37°C inkubieren.
- cDNA bei -20°C lagern.

Die Qualität der cDNA-Synthese wurde mit einer RT-PCR für das Gen *Beta-Actin* kontrolliert. Alle verwendeten Hoden-, Hirn-, Leber-, Nieren- und Milzgewebe, sowie Leukozyten und männliche und weibliche B-Zellinien wurden mit einer solchen RT-PCR überprüft (Abb. 3-1). Die *Beta-Actin* Primer sind in Tabelle A-41 aufgeführt.

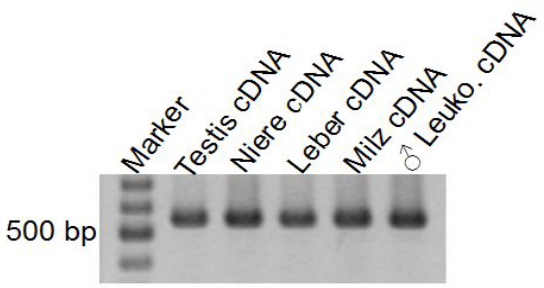


Abb. 3-1: *Beta-Actin* Kontrolle am Beispiel *Macaca mulatta*.
Beta-Actin (NM_001033084) PCR-Produkt mit P518/519 ergibt eine Bande von 535 bp. Expression in allen untersuchten Mmul Geweben vergleichbar. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

3.1.3. Amplifikation von DNA und cDNA

3.1.3.1. Standard Polymerase-Kettenreaktion

Mit Hilfe der Polymerasekettenreaktion (PCR: Polymerase chain reaction) können definierte DNA-Abschnitte aus einer geringen DNA-Ausgangsmenge *in vitro* amplifiziert werden. Für die PCR werden DNA-Fragmente, die bekannte Sequenzen im 5´- und im 3´- Bereich aufweisen, benötigt, um aus diesem Bereich Oligonukleotide (Primer) ableiten zu können. Das Grundprinzip der PCR unterteilt sich in drei Reaktionsschritte, die zyklisch 25-35 Mal

durchlaufen werden. Zuerst wird die doppelsträngige DNA bei 95°C denaturiert. Im zweiten Schritt lagern sich die Primer an die DNA (Annealing).

Dies geschieht meistens bei Temperaturen im Bereich von 50°C bis 65°C (die Temperatur hängt von der Länge der Primer und ihrer Basenzusammensetzung ab). Im dritten Schritt beginnt die Polymerisationsreaktion, bei der die Primer als Startpunkt dienen. Die Polymerase heftet sich an die angelagerten Primer und synthetisiert bei 72°C einen komplementären Strang (Extension). Dies erfolgt immer in 3´-Richtung. Bei jedem neuen Zyklus dient die neusynthetisierte DNA ebenfalls als Matrize. Somit erfolgt schließlich die exponentielle Vermehrung der gewünschten DNA-Sequenz.

Mit der Entdeckung der *Taq*-Polymerase des thermophilen Bakteriums *Thermus aquaticus* wurde die PCR als Standardmethode etabliert. Die *Taq*-Polymerase hat ihr Aktivitätsoptimum bei 72°C. Sie toleriert jedoch auch Temperaturen von 95°C, welche beim Denaturieren der DNA benötigt werden.

Reaktionsansatz mit cDNA für 25 µl (Invitrogen *Taq*):

- 1 µl cDNA
- 0,25 µl Forwardprimer
- 0,25 µl Reverseprimer
- 0,25 µl dNTP-Mix
- 3,75 µl 10x Puffer Ø Mg^{++}
- 1,25 µl $MgCl_2$
- 0,3 µl *Taq*-Polymerase (5U/µl)
- 17,95 µl steriles H_2O

PCR-Bedingungen für RT-PCR:

95°C – ∞
95°C – 3 min
95°C – 30 sec ┐
61°C – 45 sec │ 35 x
72°C – x min ┘
72°C – 5 min
4°C - ∞

Je nach Einsatz der PCR, bzw. RT-PCR wurden in dieser Arbeit 25µl, oder 50µl PCR Ansätze gewählt. Je nach Beschaffenheit der DNA (z.B. C,G-Gehalt) und Verwendung der amplifizierten DNA wurden unterschiedliche Polymerasen eingesetzt. Für die Amplifizierung der *DDX3X* Promotersequenzen wurde die Promega Go*Taq* eingesetzt, da sie eine effiziente Amplifikation in (C,G)-reichen Sequenzen erlaubt. 5´UTR-Sequenzen, für die *in-vitro* Translation, und die Mutagenese-PCR wurden mit einer *Pfu*-Polymerase durchgeführt. Die RT-PCRs zur Kartierung der TSSs wurden mit der Invitrogen *Taq* durchgeführt. Die Extensionszeit x richtet sich nach der Länge der zu generierenden PCR-Produkte. Hierbei wurde für jedes kb des zu amplifizierenden

DNA-Fragments 1 min (mit *Taq* Polymerase) für die Extensionszeit angesetzt. Für PCRs mit *Pfu*-Polymerase wurden 2 min Extensionszeit pro kb DNA-Fragment verwende, da die *Pfu*-Polymerase, auf Grund der Proofreading-Aktivität, weniger prozessiv ist. Als weitere Anwendung der PCR-Methode wurden auch Kolonie-PCRs, mit nicht gereinigter Plasmid-DNA, zur schnellen Kontrolle der klonierten Plasmide durchgeführt. Hierzu wurde mit einer Impfschlinge ein Teil einer transformierten Bakterienkolonie von der LB-Agarplatte abgenommen und in Wasser gelöst. 5µl dieser Bakteriensuspension wurden für eine PCR im 50µl Maßstab eingesetzt.

3.1.3.2. Schnittstellengenerierung mit Hilfe modifizierter Oligonukleotide

Um DNA-Fragmente gerichtet oder ungerichtet in Vektoren zu inserieren, wurden Oligonukleotide synthetisiert, die an ihren 5´-Enden Schnittstellen für Restriktionsenzyme aufweisen. Diese Oligonukleotide wurden so hergestellt, dass im 3´- Bereich 18 bis 22 bp komplementär zu dem zu amplifizierenden DNA-Fragment waren und im 5´-Bereich 10 bis 12 Basenpaare, die die Schnittstelle repräsentierten. Dem nichtkomplementären Bereich wurden – zusätzlich zu den Erkennungssequenzen der Restriktionsendonukleasen – Linker-Nukleotide angefügt, da die meisten Enzyme neben ihrer Konsensussequenz noch einige flankierende Basen benötigen, um ihr spezifisches Motiv zu erkennen. Die Länge dieser Abschnitte unterscheidet sich von Enzym zu Enzym. Da durch diesen nichtkomplementären 5´-Bereich die Hybridisierungsbedingungen der Oligonukleotide nicht optimal sind, wurden anfänglich 5 Amplifikationszyklen mit niedrigerer unspezifischer Anlagerungstemperatur von 55°C gefahren. Nachfolgend wurden 30 Zyklen bei einer spezifischen Anlagerungstemperatur von 60°C durchgeführt. Dieses Verfahren wurde für die Klonierung von *DDX3X* Promotersequenzen, sowie von (C,G)-reichen Sequenzen, angewendet und nach folgendem Schema durchgeführt:

PCR-Bedingungen:

95°C – ∞
95°C – 5 min
95°C – 30 sec ┐
55°C – 30 sec │
72°C – x sec ┘
95°C – 30 sec ┐
60°C – 30 sec │ 30x
72°C – x sec ┘
72°C – 10 min
4°C - ∞

3.1.3.3. Mutationsgenerierung mit Hilfe modifizierter Oligonukleotide

Die Einführung von Mutationen in eine Vektor-DNA erfolgte nach dem Stratagene-Protokoll für die „QuickChange site-directed muatgenesis“. Die Primer für die gezielte Mutation von DNA-

Sequenzen wurden mit dem Programm PrimerX gestaltet. Die Methode beruht darauf, daß in *E. coli* Bakterien synthetisierte Plasmide durch die „dam Methylase“ methyliert werden. Diese parentalen Plasmide dienen als Matrize in der PCR-Reaktion. Für diese Reaktion wird die Proofreading Polymerase *Pfu* benötigt, um eine korrekte Plasmid-Neusynthese zu gewährleisten. Nach Aufreinigung der PCR werden die parentalen Plasmide mit dem Restriktionsenzym *DpnI* geschnitten. Die *in-vitro* neu synthetisierten Plasmide tragen die über die Primer eingeführte Mutation und können nachfolgend in kompetente *E. coli* Zellen transformiert und propagiert werden.

Reaktionsansatz:

- 50 - 100ng gereinigte Plasmid-DNA
- 1 µl Forwardprimer (mit Mismatch)
- 1 µl Reverseprimer (mit Mismatch)
- 1 µl dNTP-Mix
- 5 µl 10x Puffer
- 1 µl *Pfu*-Polymerase
- 5 µl DMSO
- Auf 50 µl mit H_2O auffüllen

PCR-Bedingungen:

95°C – 5 min
95°C – 30 sec
95°C – 30 sec ┐
50 - 55°C – 1 min ├ 18x
68°C – 14 min ┘
4°C - ∞

- Aufreinigung zur Entfernung von DMSO mit „QIAquick Gel Extraction Kit“
- 1 h Verdau mit *DpnI*

3.1.4. Klonierungstechniken

3.1.4.1 Restriktionsspaltung doppelsträngiger DNA

Restriktionsendonukleasen sind Enzyme, die bestimmte Basenabfolgen auf doppelsträngiger DNA (dsDNA) erkennen, und diese spalten. In dieser Arbeit wurden Restriktionsendonukleasen vom Typ II eingesetzt, bei denen Erkennungssequenz und Schnittstelle identisch sind. Die Sequenzspezifität der Restriktionsendonukleasen hängt entscheidend von der Ionenkonzentration in der Reaktionslösung sowie von der Inkubationstemperatur ab. Daher erfolgte die Spaltung der DNA jeweils bei der vom Hersteller empfohlenen Temperatur und in dem entsprechenden Reaktionspuffer für eine Stunde.

3.1.4.2. Dephosphorylierung von Vektoren

Um bei einer Ligationsreaktion die Selbstligation des Vektors zu verhindern, erfolgte eine Dephosphorylierung der beiden Enden mit mit „Antarctic Phosphatase“ (AnP), da nach einem Restriktionsenzymverdau Phosphatreste an den 5’-Enden der Vektoren zurückbleiben.

Ansatz: 20 µl Restriktionsverdauansatz
5 µl 10x Reaktionspuffer
1 µl AnP (5U/µl)
24 µl steriles H_20

Die Reaktion wurde für 30 Minuten bei 37°C inkubiert. Anschließend wurde das Enzym bei 85°C über 15 Minuten inaktiviert. Vor einer weiteren Verwendung der DNA erfolgte eine Aufreinigung des Ansatzes mit dem Qiagen Gelextraktionskit.

3.1.4.4. Ligation

Die Ligation eines Vektors mit einem DNA-Fragment wird durch die *T4* DNA-Ligase katalysiert, wobei zu beachten ist, dass sich ein Überschuß an *T4* DNA-Ligase negativ auf die Ligationseffizienz auswirkt. Dieses Enzym knüpft unter ATP-Verbrauch eine 3´-OH-Gruppe kovalent mit einer 5´-Phosphatgruppe unter Ausbildung einer Phosphodiesterbindung. Für den Einbau eines DNA-Fragments (Insert) in einen Vektor wurden ein linearisierter und dephosphorylierter Vektor und ein DNA-Fragment in einem Gesamtvolumen von 20µl und einem molaren Vektor/Insert-Verhältnis von 1:3 inkubiert. Als Vektormasse wurde standardmäßig 100ng gewählt. Bei sehr (C,G)-reichen Inserts wurde dieses vor der Ligationsreaktion für 10 min bei 60°C inkubiert. Zur Berechnung der einzusetzenden Insertmasse wurde folgende Formel angewandt:

$\text{Masse}_{\text{Fragment}}$ [ng] = $\text{Länge}_{\text{Fragment}}$ [bp]/ $\text{Länge}_{\text{Vektor}}$ [bp] x 3 x $\text{Masse}_{\text{Vektor}}$ [ng]

Ligationsansatz: 100 ng Vektor-DNA
X ng Insert-DNA
2 µl 10x Ligationspuffer
0,5 µl *T4* DNA-Ligase (400U/µl)
auf 20 µl mit sterilem H_2O auffüllen.

Die Inkubation erfolgte für eine Stunde bei Raumtemperatur. Die Ligationsreaktion bei Inserts mit zwei unterschiedlichen Schnittstellen wurde über Nacht in einem Eiswasserbad durchgeführt. Da sich die Wassertemperatur kontinuierlich erhöht, wird meist die optimale Ligationstemperatur beider Schnittstellen während der Reaktion erreicht. Die T4 DNA-Ligase wurde im Anschluß bei 65°C für 10 Minuten inaktiviert, um die Transformationseffizienz zu erhöhen.

Eine besondere Form der Ligation stellt die TA-Ligation in den offenen „pGEM-T Easy“ Vektor dar. Dieser Vektor ist durch Restriktionsenzymverdau mit *EcoRV* geöffnet. An die beiden Enden wird 3´ jeweils ein einzelnes Thymin angehängt. Dadurch ist eine direkte Ligation von PCR-Produkten, die mit *Taq*-Polymerasen erzeugt wurden, möglich. Diese generieren ein Adenin am 5´Ende des PCR-Produktes. Für PCR-Produkte, die mit Pfu-Polymerasen erzeugt wurden, muß vor der TA-Ligation erst eine A-Tailing Reaktion durchgeführt werden.

A-Tailing Reaktion:

- Aufreinigung des PCR-Produktes
- 5,5 µl PCR-Produkt
- 1 µl 10x *Taq*-Puffer
- 1 µl *Taq* (5 U/µl)
- 2 µl dATPs (0,2 mM)
- 0,5 µl MgCl
- 30 min bei 70°C inkubieren.

Zur Religation von Plasmiden, die mit zwei unterschiedlichen Restriktionsenzymen geöffnet wurden, muß eine Blunt-end Ligation durchgeführt werden.

Blunt-end Reaktion:

- 14 µl geöffneter Vektor
- 4 µl T4-DNA Polymerase Puffer
- 1 U T4-DNA Polymerase
- 1 µl 2 mM dNTP-Mix
- 5 min bei Raumtemperatur inkubieren
- Zur Inaktivierung 10 min bei 70°C inkubieren.
- Nach der Blunt-end Ligation kann der Vektor in einer normalen Liagationsreaktion mit T4-DNA Ligase religiert werden.

3.1.4.5. Transformation kompetenter *E. coli*-Zellen

Um Fremd-DNA in Bakterien zu amplifizieren wurden kompetente *E. coli* Bakterienzellen transformiert. Hierzu wurde Fremd-DNA mit einem Selektionsvorteil für eine transformierte Zelle, wie zum Beispiel einem Antibiotikumresistenzgen, in die Bakterien eingebracht. Nach erfolgreicher Transformation können die Bakterien unter geeignetem Selektionsdruck inkubiert und Zellen mit Fremd-DNA selektioniert werden. Die in dieser Arbeit verwendeten chemo-kompetenten Zellen (DH5α, JM109) wurden nach Beschreibung der Hersteller transformiert (Invitrogen, Promega).

3.1.4.6. Präparation von Plasmid-DNA aus Bakterien

Diese Methode macht sich zu Nutze, dass sich Plasmid-DNA und genomische bakterielle DNA unter alkalischem pH unterschiedlich verhalten. Aufgrund der superhelikalen Struktur denaturiert Plasmid-DNA später als die genomische bakterielle DNA der Wirtszelle. Bei der anschließenden Renaturierungsreaktion bildet die genomische DNA mit den Proteinen und zellulären Bestandteilen eine unlösliche Struktur, die sedimentiert werden kann. Die Plasmid-DNA wird danach mittels Ethanol präzipitiert und in einem geeigneten Volumen aufgenommen. Für Anwendungen die sehr reine DNA erfordern, wie z.B. die Transfektion von eukaryoten Zellen, wurde das Fertigsystem „QIAprep Spin Miniprep Kit“ der Firma Qiagen, nach Anleitung des Herstellers, eingesetzt. Die Plasmid-Präparation wurde im Mini-Maßstab durchgeführt. Hierzu wurden die Bakterien in 5 ml LB-Medium mit einem geeigneten Antibiotikum für 16 h bei 37°C inkubiert und im Anschluss sedimentiert.

Die Bakterien wurden am Folgetag bei ca. 2700 x g für 10 Minuten sedimentiert. Das Pellet wurde in 250µl kaltem (4°C) Puffer P1 resuspendiert und in ein Reaktionsgefäß überführt. Die Lyse der Bakterien erfolgte, nach Zugabe von 250µl Puffer P2, bei 5 min Raumtemperatur. Die Neutralisation erfolgte anschließend mit 350µl Puffer P3. Nach mehrmaligem Invertieren wurde bei maximaler Umdrehungszahl für 10 Minuten zentrifugiert. 900µl des klaren Überstandes wurden in ein neues 1,5 ml Reaktionsgefäß überführt. Die Weitere Aufreinigung der Plasmid-DNA erfolgt mit Isopropanol (Puffer PB) und Ethanol (Puffer PE) beinhaltenden Puffern. Hierzu werden die im Kit enthaltenden Säulen genutzt und die Puffer durch Zentrifugation durch die Membran befördert. Das Ethanol wurde vor Elution mit sterilem H_2O (30 bis 50µl), vollständig abzentrifugiert.

3.1.4.7. Lagerung von Bakterienkulturen

Bakterienkulturen können in LB-Medium, mit Glycerin versetzt, bei -80°C langfristig gelagert werden. Hierzu wurden 500µl einer Über-Nacht-Kultur in LB-Medium mit 500µl Glycerin in ein 1,5 ml Reaktionsgefäß überführt. Die Probe wird, bis zur vollständigen Durchmischung, im Vortex-Schüttler gehalten.

3.1.5. Ethanolpräzipitation von Nukleinsäuren

Zur Präzipitation und Konzentrierung von DNA wurde eine DNA-Lösung mit 1/10 Vol. 3 M Natriumacetat Lösung (pH 7,6) und 2,5 Vol. EtOH abs. versetzt, gemischt und für 30 min bei -20°C inkubiert. Die anschließende Sedimentation der DNA erfolgte durch 30 min Zentrifugation bei 14.000 x g. Die pelettierte DNA wurde 1x mit 70%igem Ethanol gewaschen, kurz getrocknet und anschließend in H_2O aufgenommen.

3.1.6. Konzentrationsbestimmung von Nukleinsäuren durch Messung der Optischen Dichte

Konzentration und Reinheit einer DNA-, oder RNA-Lösung kann durch Bestimmung der Extinktion in einem Photometer bestimmt werden. Doppelsträngige DNA zeigt im Spektralphotometer ein Absorptionsmaximum bei 260 nm (OD_{260nm}). Hierbei entspricht eine Absorption von 1 einer DNA-Konzentration von 50 µg/ml. Einzelsträngige DNA hat bei einer Absorption von 1 eine Konzentration von 33 µg/ml und RNA von 40 µg/ml.

Um eine eventuelle Proteinverunreinigung der Probe zu bestimmen, wird der Proteingehalt bei einem Absorptionsmaximum von 280 nm gemessen. Liegt das Verhältnis von DNA/Protein-Konzentration (Ratio A_{260}/A_{280}) bei 1,8 bis 2,0, ist die DNA-Probe weitgehend frei von Proteinverunreinigungen. Reine RNA-Proben sollten einen Wert von ~ 2,0 (Ratio A_{260}/A_{280}) aufweisen. Werte unter 1,7 zeigen eine Verunreinigung der Probe mit Proteinen an, da erhöhte Absorptionswerte bei 280 nm gemessen werden. Zusätzlich wird auch das Verhältnis A_{260}/A_{230} gemessen, welches die Reinheit von Salzen, Phenolen und Kohlenhydraten angibt. Der Wert A_{260}/A_{230} sollte bei ~ 2,0 liegen.

3.1.7. DNA-Sequenzierung – nach Kettenabbruchverfahren

Die eingesetzte Sequenzierungsmethode beruht auf dem Abbruch-Verfahren nach Sanger et al. (1977). Das Grundprinzip der enzymatischen Reaktion ähnelt dabei dem der PCR. Auch hier wird sich die Fähigkeit der Polymerase zunutze gemacht, einen neuen DNA-Strang komplementär zu einem einzelsträngigen Template aufzubauen., so dass verschieden lange Fragmente in nachweisbarer Menge entstehen. Durch den Einsatz von modifizierten Nukleotiden kommt es statistisch an jeder Nukleotidposition der Ziel-DNA zu einem Kettenabbruch, und damit zur Synthese von unterschiedlich langen Fragmenten. Die modifizierten Nukleotide enthalten 2´, 3´- Didesoxyribosen (ddNTP), wodurch die 5´→3´ Verknüpfung der DNA-Bausteine durch das Fehlen der OH-Gruppe am 3´-Ende unterbrochen wird. Werden Primer, einzelsträngiges Template, Nukleotide und die modifizierten Nukleotide (z.B. ddATP) in einem Reaktionsansatz vereinigt, besitzen die neu synthetisierten Sequenzen zwar unterschiedliche Längen, zeichnen sich aber durch das gemeinsame Merkmal aus mit derselben Base zu enden (Adenosin) (siehe Abb. 3-2).

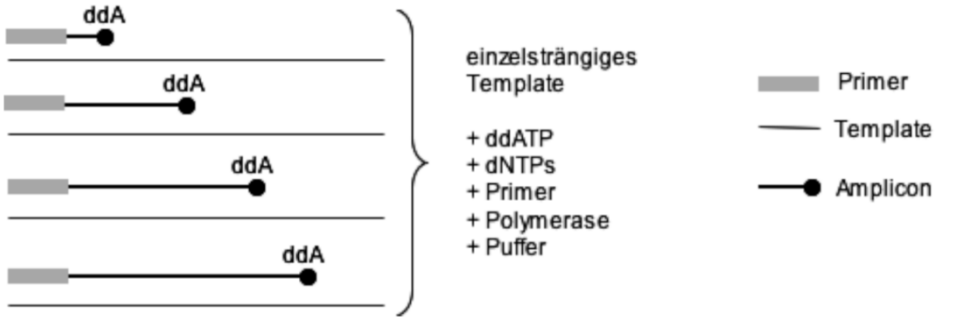


Abb. 3-2: Prinzip der Sequenzierung (nach Sanger et al., 1977).
Muster einer Ketten-abbruchreaktion am Beispiel von Amplicons mit ddA-Enden.

Um eine Aussage über die gesuchte Sequenz machen zu können, muss diese Abbruchreaktion für alle vier Nukleotide durchgeführt und analysiert werden. Durch das Durchführen einer elektrophoretischen Größenauftrennung können dann die Fragmente unterschieden und ein Rückschluss auf die Basenabfolge der Template-Sequenz gezogen werden (siehe Abb. 3-3).

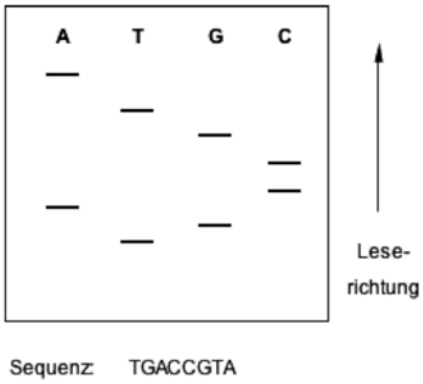


Abb. 3-3: Mögliches Bandenmuster eines sequenzierten Fragmentes.

Die Abbruchreaktion wurde mit den vier ddNTPs durchgeführt und in vier Spuren aufgetragen.

Durch den Ersatz einer thermostabilen Polymerase lässt sich die Cycle-Sequenzierung durchführen. Die Synthese erfolgt dann, ähnlich wie bei der PCR, durch die Abfolge der drei Phasen: Denaturierung, Annealing und Extension. Der Unterschied einer Direktsequenzierung von PCR-Produkten zur PCR ist jedoch, dass kein exponentielles Wachstum stattfindet, sondern lediglich eine Verdopplung einer Sequenz erfolgt. Es wird nur einer der beiden DNA-Stränge durch den Einsatz eines Primers amplifiziert. Eine Auswertung der Ergebnisse wäre sonst nicht möglich. Über den Einsatz von vier Fluoreszenz-markierten Nukleotiden ist es möglich, den Reaktionsansatz in einem Reaktionsgefäß durchzuführen und die Fragmente gleichzeitig zu bestimmen (Abb. 3-4).

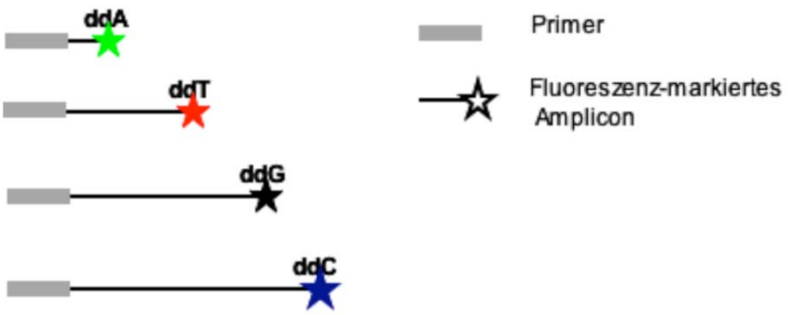


Abb. 3-4: Prinzip der Sequenzierung mit Fluoreszenz-markierten ddNTPs (nach Sanger et al., 1977).

Mögliches Muster einer Ketten-abbruchreaktion.

Der eingesetzte ABI PRISM 3100 Genetic Analyzer ist ein Fluoreszenz-Detektions-System, bei dem die Größenauftrennung der Nukleinsäure-Fragmente durch Kapillarelektrophorese erfolgt. Die Fragmente werden auf eine mit Polymeren (POP4) befüllte Kapillare gegeben und, ähnlich wie bei der Agarosegel-Elektrophorese, der Größe nach aufgetrennt. Ein Argonlaser ist dabei auf ein Fenster am Ende der Kapillare gerichtet und regt die Fluoreszenz-Farbstoffe zu spezifischen Emissionen an. Diese Emissionen werden durch einen Chip in einer CCD-Kamera (*charge coupled device*-Kamera) detektiert und durch eine Rechnereinheit in digitale Daten umgewandelt. Der detaillierte Aufbau und die Funktionsweise des ABI-3100-Systems sind im

Handbuch von „Applied Biosystems“ ausführlich beschrieben. Der Vorteil dieses Systems ist die kurze Laufzeit der Sequenzierung, bei der 16 Proben parallel analysiert werden können.

3.1.7.1. Ansatz der Sequenzreaktion

Die Sequenzierungsreaktion wurde nach dem in Tabelle 3-1 aufgeführten Schema angesetzt. Das durchgeführte PCR-Programm für die Reaktion ist in Tabelle 3-2 dargestellt.

Tab. 3-1: Pipettierschema für die Sequenzierungsreaktion der pGL4-Konstrukte.

Volumen [µl]	**Reagenz**
2	BigDye Ready Reaction Premix
1	Primer
1 – 5	DNA-Template, z.B. Vektor (200 – 500ng)
Auf 10 µl auffüllen	steriles H_2O
10	**Total**

Tab. 3-2: PCR-Programm für die Sequenzierungsreaktion mit dem „BigDye Terminator v1.1 Cycle Sequencing Kit“ (Protocol No. 4337036A, Applied Biosystems).

Die Schritte 2 bis 4 unterliefen insgesamt 25 Zyklen.

Schritt	**Temperatur [°C]**	**Zeit**
1	95	5 min
2	95	20 sec
3	52	15 sec
4	60	45 sec
5	60	4 min
6	4	∞

3.1.8. Agarose-Gelelektrophorese

DNA-Fragmente können entsprechend ihrer Größe im elektrischen Feld aufgetrennt werden. Da die Phosphatreste der DNA dem Molekül eine negative Ladung verleihen und mit einem basischen Laufpuffer (1xTAE) gearbeitet wird, wandert die DNA in Richtung der Anode (Pluspol). Die Laufeigenschaften im Agarosegel werden vom Molekulargewicht, von der Konformation der DNA, der Höhe der angelegten Spannung sowie der Konzentration des Gels beeinflusst.

Als konventionelle Gelelektrophorese bezeichnet man die Auftrennung von DNA-Fragmenten zwischen 70 bp und 60 kb in Agarosegelen, an die ein homogenes elektrisches Feld angelegt wird. Die Trennung kleiner Fragmente von 100-400 bp erfolgte in 2%igen Agarosegelen, an die eine Spannung von 6-8 V/cm angelegt wurde. Fragmente mit einer Größe von 0,5 bis 5 kb wurden in 0,8 und 1%igen Gelen aufgetrennt. Die angelegte Spannungen und die Laufzeiten der Gele variierten je nach Verwendungszweck von 2-6 V/cm und 30 min - 20 h.

Als Laufpuffer wurde 1 x TAE verwendet. Die Proben wurden zum Auftrag in 6x Gelladepuffer aufgenommen. Das im Ladepuffer enthaltene Bromphenolblau ermöglicht ein Abschätzen der Laufstrecke im Gel, während Ficoll ein Absinken der DNA in die Geltaschen sicherstellt.

Zur Anfärbung der DNA enthielten die Gele 0,5 µg/ml Ethidiumbromid, das sich interkalierend in doppelsträngige DNA einlagert. Durch UV-Licht angeregt, fluoresziert das interkalierte Ethidiumbromid und markiert somit die Lage der DNA-Banden. Die Intensität der Fluoreszenz kann - im Vergleich zu einer Probe bekannter Konzentration - Aufschluss über die DNA-Konzentration der jeweiligen Bande geben. Zur Abschätzung der Größe der im Gel aufgetrennten DNA-Moleküle dienen Größenstandards definierter Konzentration.

3.1.9. Photodokumentation der Agarosegele

Die Zugabe von Ethidiumbromid zu Agarosegelen ermöglicht den Nachweis von DNA. Zur Dokumentation wurden die Agarosegele bei UV-Licht von 312 nm fotografiert.

3.1.10. DNA-Extraktion aus Agarosegelen

Um DNA-Fragmente als Ligationsinsert zu erhalten, mussten diese präparativ aus dem Agarosegel isoliert werden. Dazu wurde die entsprechende DNA-Bande unter UV-Licht von 312 nm vorsichtig mit dem Skalpell aus dem Gel ausgeschnitten und mit dem „Gel Extraction Kit“ der Firma Qiagen aufgereinigt.

Zur DNA-Extraktion wurde das Agarose-Stück zuerst mit dem QG-Puffer bei 60°C verflüssigt (10 min) und auf eine Säule aufgetragen. Bei hoher Salzkonzentration bindet die DNA an die Säulenmatrix. Durch Waschen mit einer Lösung niedriger Salzkonzentration (TE-Puffer oder H_2O) kann die DNA anschließend eluiert werden. Die Durchführung erfolgte nach Anleitung des Herstellers.

3.1.11. *In-vitro* Transkription/Translation von *DDX3X* 5´UTRs

Das „TNT® Quick Coupled Transcription/Translation Systems“ basiert auf dem „TNT® Coupled Reticulocyte Lysate Systems“ (Promega), welches alle benötigten Komponenten zur eukaryotischen Translation in einem einzigen Master-Mix kombiniert. Es sind die RNA-Polymerase (T7), Nukleotide, Salze, rekombinanter RNasin® Ribonuklease-Inhibitor, sowie die Retikulozyten-Lysat Lösung (Retikulozyten = unreife Vorstufen der Erythrozyten) – bestehend aus tRNAs, Aminosäuren, Initiations-, Elongations- und Terminations-Faktoren – im Master-Mix enthalten.

Als Grundgerüst für das Vektor-basierte *in-vitro* Transkriptions-/Translations-System dient der Vektor pUC18 (Invitrogen, # 15363013). Der erfolgreiche Einsatz eines Vektor-basierten *in-vitro* Systems wurde für die 5´UTR des *NRAS* Gens demonstriert (Kumari et al., 2007).

Als Vektorgerüst für die Experimente wurde der pUC18-Vektor gewählt, da der pUC18-Vektor über ein einfaches und durch die umfangreiche „Multiple Cloning Site“ (MCS) leicht anpassbares Vektorgerüst verfügt. Das pUC18-Plasmid ist ein sogenanntes „high-copy“

Plasmid, was erlaubt, bis zu 500 Kopien pro *E. coli*-Zelle zu synthetisieren. Die Möglichkeit Schnittstellen für die einfach zu handhabenden Restriktionsenzyme *HindIII* und *EcoRI* zu Anfang und Ende der MCS zu haben, ist ein weiterer Vorteil dieses vergleichsweise kleinen Vektors.

Der kommerzielle Expressionsvektor pTNT (Promega, #L5610) verfügt über *XhoI* und *NotI* Schnittstellen am Anfang und Ende der MCS. Besonders von *NotI* ist bekannt, daß die Restriktionseffizienz sehr gering ist, wenn nicht zusätzlich lange adäquate Sequenzen die Schnittstelle flankieren, was gegen die geplante Verwendung in Primersequenzen spricht. Weiterhin ist in diesem Vektor vor der MCS die 5´ Leadersequenz des β-Globin-Gens vorhanden, welche nicht ohne Mutation entfernt werden kann, und nicht den natürlichen Kontext der *DDX3X* 5´UTRs widerspiegelt. Zudem ist kein *LacZ* Gen zur Blau-Weiß-Selektion im pTNT Vektor vorhanden. Der im *in-vitro* Translationskit enthaltene Kontrollvektor (T7-Luziferase Kontroll-DNA) stellt, auf Grund der mit drei Restriktionsenzymen sehr limitierten MCS, ebenfalls keinen adäquaten Vektor dar (Abb. 3-5).

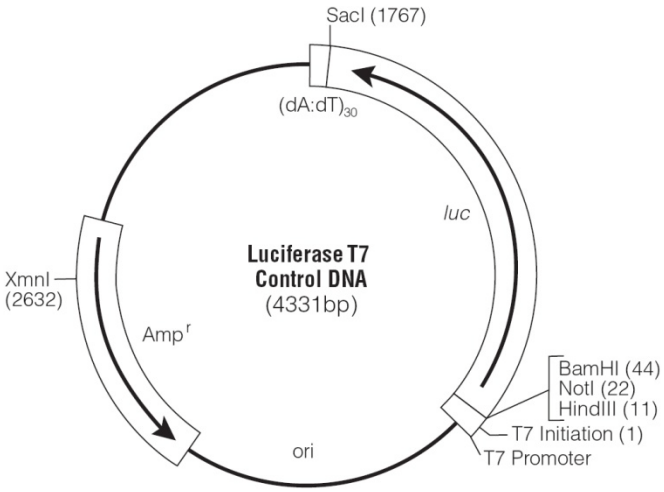


Abb. 3-5: Luziferase T7-Kontrollplasmid (Promega, # L4821).

Das für diese Arbeit gewählte Reportergen Luziferase (*Luc2*) mußte erst in den pUC18-Vektor eingefügt werden. Die verwendeten Primersequenzen sind im Anhang aufgeführt (Tab. A-44). Zum Klonieren der benötigten Sequenzen in das pUC18-Plasmid, werden die Restriktionsenzym-Schnittstellen *HindIII* an Position 400, *KpnI* an Pos. 443 und *Eco*RI an Pos. 451 genutzt (Abb. 3-6). Die restlichen acht Restriktionsenzymschnittstellen der original pUC18 MCS werden dadurch eliminiert, wodurch keine unnötigen Sequenzen in die 5´UTR-Sequenz gelangen. Die Eigenschaften des pUC18-Vektors werden für die Vervielfältigung des Plasmids in *E. coli* Zellen benötigt, dabei ist der „ori“ (Origin of Replication) für die Replikation des Plasmids zuständig und als Selektionsmarker wird das Ampicilin-Resistenzgen eingesetzt. Über die zusätzliche Blau-Weiß-Selektion des Gens *LacZ* (β-Galaktosidase) können erfolgreich transformierte Plasmide (weiße Kolonien auf LB-Agar Platten) selektioniert werden.

Die Reihenfolge der Fragmente in der klonierten Transkriptionskassette ist wie folgt: *HindIII* Schnittstelle → T7 Promotersequenz (20 bp) → *DDX3X* 5´UTR (X bp)→ *KpnI* Schnittstelle → Kozak-Sequenz GGGATGG→ *Luc2* Reportergen (1661 bp)→ Poly(A) -Sequenz (30 bp)→ *EcoRI* Schnittstelle (in Abb. 3-6 zu sehen). Die gegenläufige Orientierung der eingefügten

Transkriptionskassette zur Plasmidorientierung hat keinen Einfluß auf die Transkription der Reportergenkonstrukte, da die *in-vitro* Transkription nur von der T7-Promotersequenz abhängt.

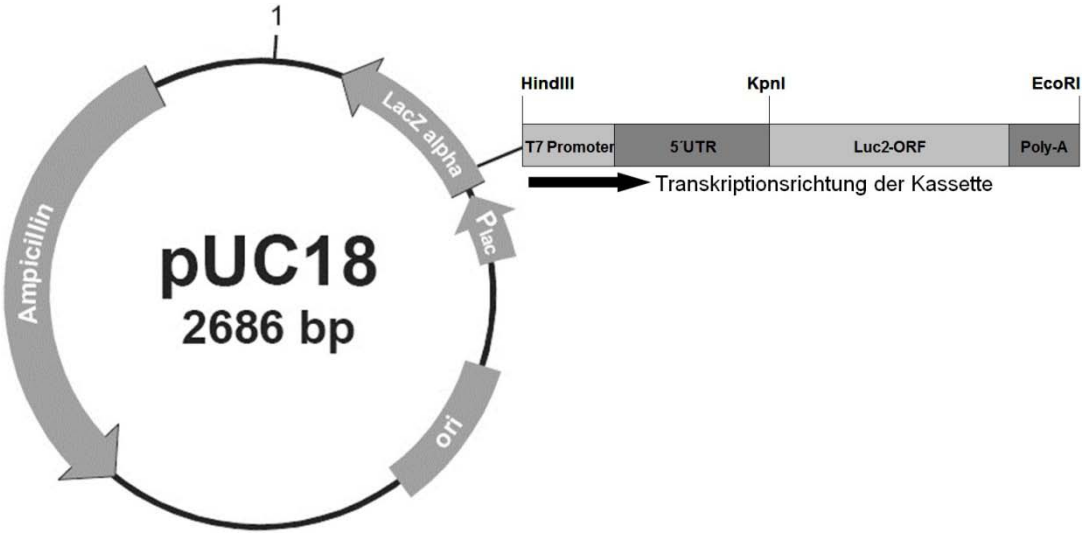


Abb. 3-6: Schematische pUC18 Vektorkarte (modifiziert nach Invitrogen).
Die klonierte Kassette ist mit den genutzten Schnittstellen *HindIII*, *KpnI* und *EcoRI* markiert. Der Pfeil gibt die Transkriptionsrichtung der Kassette an. Die T7-Promotersequenz ist jeweils in dem Forwardprimer zur Amplifikation der *DDX3X* 5´UTR-Sequenz enthalten.

In einem ersten Schritt wurde der Luziferase Leserahmen (*Luc*2), der im fertigen Vektor als Reportergen dienen soll, aus dem pGL4.10-Vektor (Promega, # E6651; Acc.No. AY738222) per PCR mit einer Proofreading Polymerase (*Pfu*-Polymerase), zur Reduzierung von Lesefehlern, amplifiziert. Der Forwardprimer ist 5´ bereits mit einer Schnittstelle für das Restriktionsenzym *KpnI* (5´…GGTACC…3´) ausgestattet und beinhaltet weiterhin die Kozak-Sequenz GGGATGG. Um einen möglichst natürlichen ATG-Kontext der *DDX3X* 5´UTR Transkripte zu erhalten wurde untenstehender Aufbau gewählt (Abb. 3-7). Das Startcodon ATG und das G an Position +4, stammen aus dem Luziferase-ORF. Die Sequenz CAGGG, unmittelbar vor dem ATG-Codon, mit dem Guanin an -3 stammt aus der *DDX3X* 5´UTR (Position 30.725 bis 30.729 im BAC RP13-13A3) und ist im Luziferase Forwardprimer integriert. Der einzige Unterschied zur *DDX3X* Kozak-Sequenz ist, daß das optimale G an +4 aus dem Luziferasegen übernommen worden ist, und nicht ein Adenin, wie es im *DDX3X* Gen an +4 vorkommt.

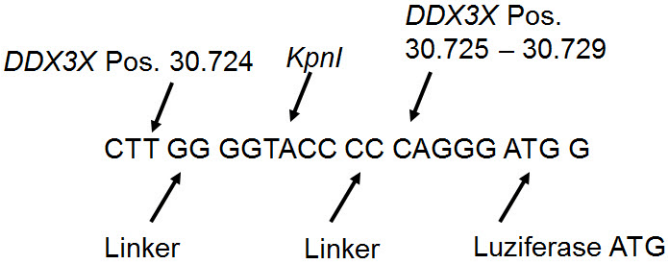


Abb. 3-7: ATG-Translationskontext der *in-vitro* pUC-18 Plasmide.
Die Verknüpfung der *DDX3X* 5´UTRs mit dem Luziferasegen erfolgt an der *KpnI* Schnittstelle. Der Reverseprimer der *DDX3X* 5´UTR enthält die Sequenzen bis einschließlich der *KpnI* Schnittstelle. Im vorbereiteten pUC18-Vektor sind die Sequenzen ab der *KpnI* Schnittstelle enthalten. Im Unterschied zu den nativen *DDX3X* 5´UTRs befinden sich in diesem Aufbau zehn zusätzliche Nukleotide stromaufwärts der Kozak-Sequenz. Positionsangaben beziehen sich auf den BAC RP13-13A3.

Im Reverseprimer für den Luziferase-ORF befinden sich, wie im pGL4.10 Vektor, zwei aufeinanderfolgende Stopcodone TAATAA und ein Poly(A)-Schwanz von 30 Nukleotiden. Der Poly(A)-Schwanz ist nicht Grundvoraussetzung für das *in-vitro* Translationssystem, stabilisiert das Transkript aber auch in einer *in-vitro* Translationsreaktion und kann somit die Translationseffizienz der Reaktion erheblich verbessern. Schließlich ist im Reverseprimer noch eine *Eco*RI Schnittstelle (5´…GAATTC…3´), zur Klonierung in den pUC18 Vektor, enthalten. Die so erhaltenen PCR-Produkte wurden zunächst in den Vektor „pGEM-T easy“ (Promega, # A1360) ligiert und zur weiteren Amplifizierung in DH5α chemo-kompetente *E. coli* Zellen transformiert. Nach Isolation der Plasmide wurden diese mit den entsprechenden Enzymen *Kpn*I und *EcoR*I in einem Restriktionsenzymverdau geschnitten. Die Reportergenfragmente wurden mittels einer Gelisolation von den pGEM-T easy Vektoren getrennt und nach Aufreinigung mit dem QIAquick Gel Extraction Kit (QIAGEN, # 28706) in den, zuvor mit *Kpn*I und *Eco*RI geöffneten und dephosphorylierten, Zielvektor pUC18 ligiert. Dieser im Folgenden „pUC18-*Luc2*“ genannte Vektor diente als Basis für alle weiteren *in-vitro* Konstrukte.

Die lange *DDX3X* 5´UTR wurde mit spezifischen unmodifizierten Forwardprimern für *DDX3X* T-TSS-III (Primer 2526) und einem spezifischen *DDX3X* Reverseprimer, der Exon1/2 (P1521) überlappend lokalisiert ist, mittels RT-PCR aus humaner Hoden cDNA amplifiziert. Die *DDX3X* 5´UTRs wurden zunächst ebenfalls in den pGEM-T easy Vektor kloniert. In einem zweiten Schritt wurden die 5´UTR-Inserts nochmals mit modifizierten Primern in einer zweiten PCR amplifiziert. Diese Forwardprimer beinhalten zusätzlich *HindIII* (5´…AAGCTT…3´) Schnittstellen und die T7-Promotersequenz. Der Reverseprimer, der in diesem Fall am 3´Ende der *DDX3X* 5´UTR Sequenz ansetzt, enthält wiederum eine *KpnI* Schnittstelle, um die *DDX3X* 5´UTRs mit dem *Luc2*-ORF im pUC18-Plasmid verbinden zu können. Die PCR-Produkte wurden erneut in pGEM-T easy Vektoren kloniert und amplifiziert. Nach Isolation der Plasmide mittels erneuter Mini-Präparation wurden die Plasmide mit den entsprechenden Enzymen *KpnI* und *HindIII* geschnitten. Die *DDX3X* 5'UTR Fragmente wurden durch Gelisolation von den pGEMT-easy Vektoren getrennt und nach Aufreinigung in den Zielvektor pUC18-*Luc2* ligiert.

Der *in-vitro* Transkriptions-/Translations-Assay wird in einem 50µl Maßstab durchgeführt. Es werden 40µl Master-Mix des „TnT Quick Coupled Transcription/Translation System“ mit 1µl Methionin, 1µg Plasmid-DNA und x µl H_2O in 50µl zusammengebracht, und für 90 min bei 30°C inkubiert. Nach 90 min Inkubation werden pro Messung 2,5µl des Ansatzes entnommen und mit 50µl des „ONE-Glo Luciferase Assay System“ (Promega, # E6110) zusammengebracht. Das Luziferase-Reagenz dient zur Detektion von Lichtquanten in einem Luminometer, die durch Umsetzung des Substrates Luziferin durch das Enzym Luziferase entstehen (Abb. 3-8). Die Höhe des gemessenen Luziferasewertes in diesem Assay, soll im Folgenden als Maß für die

Translationsfähigkeit einer mRNA dienen. Die Auswertung über ein Luziferase-System ist zudem eine gute Möglichkeit der Qualitätskontrolle, da nur Luziferasetranskripte voller Länge und somit nur vollständige Luziferase-Enzyme aktiv sind. Ein weiterer Vorteil der Luziferase ist, daß sie keine post-translationale Modifikationen für ihre Aktivität benötigt, sondern sofort einsatzfähig ist. Im Gegensatz zum „Dual-Luziferase System“ wird kein Vergleich zu einer internen Kontrolle verwendet, bei dem gegen einen Co-Reporter normalisiert wird. Die Messdauer jeder Probe beträgt zehn Sekunden. Die Negativ-Kontrolle ist ein Master-Mix ohne Einsatz von DNA.

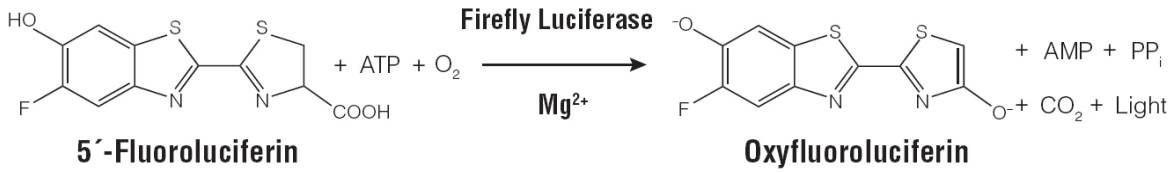


Abb. 3-8: Struktur des modifizierten Substrates 5´Fluoroluziferin aus dem ONE-GLO Luciferase Assay System“ (Promega, # E6110).

Das neu entwickelte 5´Fluoroluziferin bietet eine deutlich längere Stabilität und Haltbarkeit der Komponenten während der Reaktion und eine höhere Toleranz gegenüber chemischen Komponenten, die in Zellkultur Substraten, oder *in-vitro* Substraten vorkommen können. Die Halbwertszeit der Reaktion beträgt ca. eine Stunde. Das Reaktionsprinzip, bei welchem aus Luziferin, ATP und molekularem Sauerstoff unter Zugabe von Luziferase und Magnesiumionen Lichtquanten erzeugt werden, wurde nicht verändert.

3.2. Zellbiologische Methoden

3.2.1. Kultivierung und Lagerung von humanen Zelllinien

In dieser Arbeit wurden ausschließlich adhärente Zelllinien verwendet. Die Zellen wurden unter definierten Bedingungen in Kultur zur Proliferation angeregt und zur langfristigen Lagerung in einem entsprechenden Einfriermedium (Medium + 10 % DMSO) in flüssigem Stickstoff eingefroren.

Um die verwendeten Zellen in Kultur zu halten, mussten diese in regelmäßigen Abständen passagiert werden. Da es sich bei den verwendeten Zelllinien um adhärente Zellen handelt, mussten diese durch die Zugabe von Trypsin-EDTA von der Oberfläche gelöst werden. Hierzu wurden die Zellen zunächst mit PBS gewaschen, und nach Zugabe von Trypsin-EDTA (500 µl 10x Trypsin-EDTA + 4.500 µl PBS) für etwa 5 - 10 min bei 37°C inkubiert. Die Zellen wurden anschließend mechanisch vom Kulturflaschenboden gelöst und in ein 14 ml Falcon-Röhrchen mit Vollmedium, zum Abblocken der Trypsinierung, überführt. Die Sedimentation der Zellen erfolgte bei RT und 1.200 rpm. Das Zellpellet wurde in 5 ml Vollmedium resuspendiert und jeweils, in einem Verhältnis von 1:10, auf neue Kulturflaschen aufgeteilt.

3.2.2. Promotor- und Enhanceranalysen im zellulären System

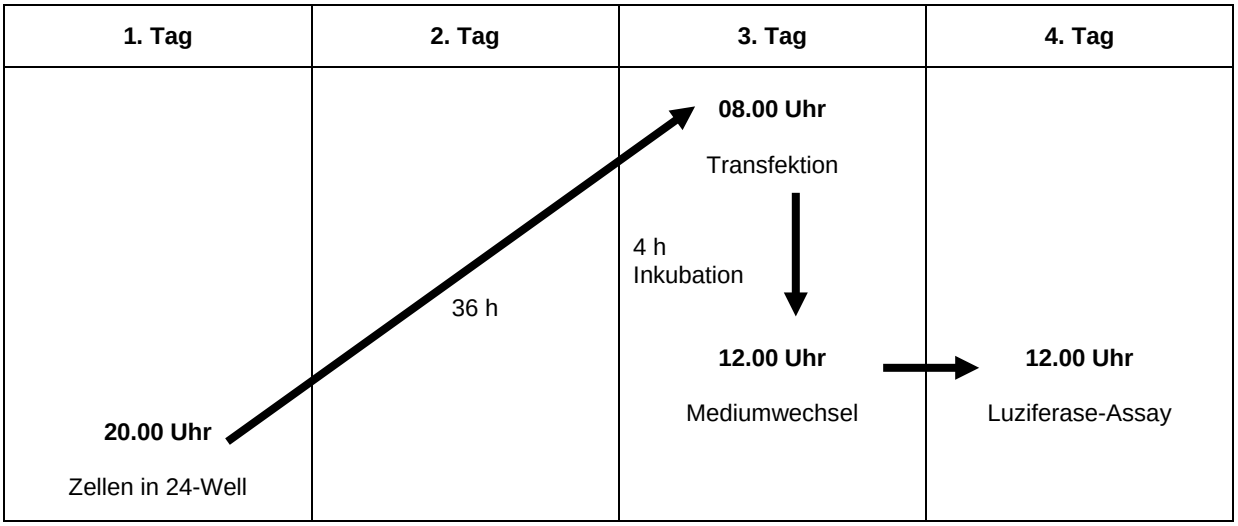
Zur Untersuchung *cis*-regulatorischer Elemente wurden Reportergenkonstrukte generiert und transient in Human EP2102 Zellen transfiziert. Das Reportergen Firefly-Luziferase stand unter der Kontrolle der zu untersuchenden *cis*-regulatorischen Elemente. Je nach Aktivierungspotential erfolgte eine stärkere oder schwächere Expression des Reportergens. Als Kontroll-Vektor diente der pGL4.13 Vektor, bei dem das Firefly-Luziferasegen unter der Kontrolle des viralen SV40-Promotors (Simian Virus 40) steht. Dieser weist in nahezu allen Zellen eine starke Transkriptionsaktivierung auf. Um die Transfektionseffizienz der einzelnen Ansätze bestimmen und vergleichen zu können, wurden Kotransfektionen mit Kontrollvektoren (pGL4.74-TK), die das Gen für die *Renilla*-Luziferase tragen, durchgeführt. Die Auswertung der Kotransfektionsexperimente wurde mit dem „Dual-Luciferase Reporter Assay System“ der Firma Promega durchgeführt.

3.2.3. Transiente Transfektion von Reportergenkonstrukten in eukaryotischen Zellen mittels Transfektion mit nicht-liposomalen Lipiden

In dieser Arbeit wurden transiente Transfektionen an EP2102-Zellen mit dem Metafectene Transfektionsreagenz durchgeführt. Hierbei handelt es sich um ein Transfektionssystem aus kationischen nicht-liposomalen Lipiden. Die DNA wird in einem ersten Schritt mit einem Enhancer-Reagenz kondensiert. In einem zweiten Schritt werden Lipide (Metafectene-Reagenz) zugegeben, die die kondensierte DNA micellenartig umhüllen. Vorteile dieser Methode sind die geringen DNA-Mengen, die benötigt werden und die geringe Toxizität der Micellen. Für die Transfektion wurden am Vortag circa 80.000 Ep2102-Zellen mit einer Konfluenz von 60-80% in „24-Loch-Kulturplatten“ ausgesät und in 0,5ml Vollmedium (incl. β - Mercaptoethanol 1 µl/ml Medium) für 36 h kultiviert, so dass am Tag der Transfektion die Kultur eine Konfluenz ~ 80% erreicht hatte. Pro Konstrukt wurden jeweils fünf Wells vorbereitet. Die Zellzahl wurde mit einer 1:1 Lösung aus Zellsuspension (10 µl) und Trypanblau (10 µl), in einer Neubauer-Zählkammer unter dem Mikroskop, bestimmt.

Die *DDX3X* Promoterkonstrukte, sowie die *MSY2* und *MSY2-X* Konstrukte sind alle in Kapitel 4.9. und 4.10 beschrieben. Alle verwendeten Primer zur Herstellung der Reportergenkonstrukte sind in der Tabelle A-42 im Anhang aufgelistet. Die vorbereitete DNA-Lösung (Mix1), bestehend aus 950ng eines pGL4-Reportergenkonstruktes, 50ng des pGL4.74 *Renilla* Co-Reporters und 50µl serumfreies/ antibiotikafreies Medium wurde mit Mix2, bestehend aus 4µl Metafectene (pro Well) und 30µl serumfreies/antibiotikafreies Medium gemischt und zur Ausbildung der Micellen für 20 min bei Raumtemperatur inkubiert. In der Zwischenzeit wurden die Zellen einmal mit PBS gewaschen und mit 0,5 ml frischem serumfreiem/antibiotikafreiem Medium versetzt. Der DNA-Mizellen-Komplex wurde gleichmäßig über die Zellen verteilt.

Nach dieser ersten Inkubationsphase wurden die Zellen nach 4 h einmal mit PBS gewaschen und mit 500µl frischem Vollmedium (incl. β-Mercaptoethanol 1µl/ml Medium) für weitere 24 Stunden inkubiert. Die Inkubation erfolgte über 28 Stunden, um genügend Luziferasemoleküle im Zellysat, für einen „Dual-Luciferase Reporter Assay“ zu erhalten. Das Verhältnis von experimentellem Reporter zu Co-Reporter betrug 20:1, wobei insgesamt 1000ng DNA pro Ansatz transfiziert wurden. Der Einsatz von geringen *Renilla* Co-Reporter Mengen ist empfohlen, um *trans*-Effekte zwischen den Promotoren der *Renilla* Vektoren und den zu untersuchenden regulativen Elementen zu vermeiden. Die „Dual-Luciferase Reporter Assays“ auf dem Luminometer MicroLumat LB96 der Firma Berthold durchgeführt. Die transiente Transfektion erfolgte nach folgendem Schema:



3.2.4. „Dual-Luciferase Reporter Assay“ Methode

Das Prinzip dieser Nachweismethode beruht auf einer simultanen Expression und Messung unterschiedlicher Reportergene. Beide Vektoren tragen die Luziferasegene verschiedener Spezies, die bei der Zugabe spezifischer Substrate Licht freisetzen. Diese Lichtemission kann von einem Luminometer detektiert werden. Die Firefly-Luziferase des Glühwürmchens *Photinus pyralis* katalysiert eine ATP- und Magnesium-abhängige, oxidative Decarboxylierung von Luciferin unter Freisetzung von Licht der Wellenlänge 562 nm (Abb. 3-8). Die Lumineszenzreaktion der *Renilla*-Luciferase der Qualle *Renilla reniformis*, die unter der Kontrolle des HSV-TK Promotors, aus dem Herpes simplex Virus steht, findet ebenfalls durch eine oxidative Decarboxylierung statt. Allerdings wird bei dieser Reaktion Coelenterazin anstelle von Luziferin, unter Freisetzung von Licht, umgesetzt (Abb. 3-9).

Die Vorteile eines „Dual-Luciferase Reporter Assays“ im Vergleich zu einem „Single-Luciferase Reporter Assay“ liegen in der Reduktion experimenteller Schwankungen, der Normalisierung systembedingter Fehler, wie z.B. Zytotoxizität der Reagenzien und der Normalisierung der Transfektionseffizienz durch Bildung eines Quotienten aus Firefly- und *Renilla*-Werten. Da eine Zelle im Verhältnis immer gleichviel Reporter- wie Co-Reporter-Vektoren aufnehmen wird, ist

der Vergleich von Messdaten verschiedener Tage, die sich in Zelldichte, Lebensfähigkeit und einer unterschiedlichen Aufnahmekapazität für Fremd-DNA unterscheiden können, möglich. Des Weiteren können die „24-Loch-Kulturplatten“ selbst Unterschiede („edge effects“) in der Kultur erzeugen, da nicht in jeder Vertiefung dieselben Bedingungen bezüglich Temperatur und Luftfeuchte herrschen können.

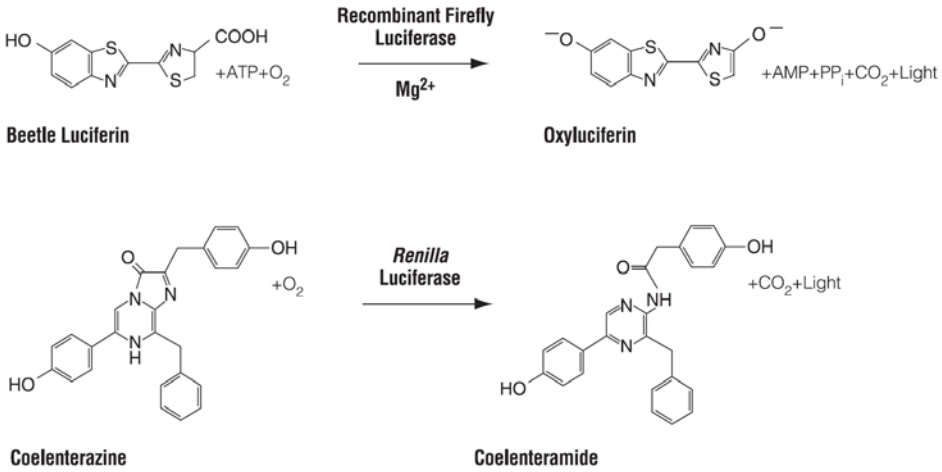


Abbildung 3-9: Von Firefly- und *Renilla*-Luziferase katalysierte Biolumineszenzreaktionen (Dual-Luciferase Reporter Assay System, Technical Manual # 040, Promega).
Beide Luziferasen sind monomere Proteine mit einem Molekulargewicht von 61kDa für die Firefly- bzw. von 36 kDa für die *Renilla*-Luziferase.

Alle Arbeiten mit diesem System wurden nach Anleitung des Herstellers des „Dual-Luciferase Reporter Assays“ (Promega, Mannheim) durchgeführt. Hierzu wurden die Zellen einmal mit PBS gewaschen und anschließend mit 100µl „Passive Lysis Buffer“ (PLB) eingedeckt. Die Lyse der Zellen erfolgte für 30 min auf einem Orbitalschüttler. 20µl Lysat wurden aus jeder Vertiefung der „24-Loch Kulturplatten“ in eine geeignete „96-Loch-Platte“ überführt. In einer ersten Messung wird die Aktivität der Firefly-Luziferase für 10s nach Injektion von 100µl „Luciferase Assay Reagent II“ (LARII-Puffer) in einem Luminometer bestimmt. In einem zweiten Schritt wird die erste Reaktion durch Zugabe von 100µl des zweiten Substrats „Stop & Glo“ inhibiert, und die *Renilla* Reaktion wird in Gang gesetzt. Der zweite Messdurchlauf erfolgt unter denselben Bedingungen. Für jedes Konstrukt wurden je fünf EP2102 Wells transfiziert. Aus fünf Messungen für ein Konstrukt wurden jeweils die zwei extremsten Werte nicht verwendet. Die relativen Licht-Einheiten (RLU) der Firefly-Luziferase, dividiert durch die RLUs der *Renilla*-Luziferase ergaben ein auf die Menge an transfizierten Zellen normiertes Ergebnis. Der Firefly zu *Renilla* RLU-Quotient der experimentellen Konstrukte wurde aus den Mittelwerten von drei Messungen gebildet und anschließend mit dem RLU-Quotient der Positivkontrolle pGL4.13 verglichen. Die Positivkontrolle wurde standardmäßig als 100% gesetzt. Jedes Transfektions-experiment wurde mindestens an zwei unabhängigen Tagen wiederholt.

4. Ergebnisse

4.1. Keimbahn-Komplexität der *DDX3Y* Transkripte und Kartierung der dazugehörigen Core-Promotermotive

Im Rahmen dieser Doktorarbeit wurde versucht, ein möglichst vollständiges Bild der verschiedenen *DDX3Y* und *DDX3X* Transkriptklassen in Humangeweben, durch die Kartierung der dazugehörigen Transkriptionsstartstellen (TSSs), zu erlangen.

Um dieses Ziel zu erreichen wurden die in unserem Labor über 5´RACE und RT-PCR Experimente gewonnenen Ergebnisse mit den Daten der DBTSS-Datenbank („Database of Transcriptional Start Sites“) verglichen (Wakaguri et al., 2008). Die DBTSS-Datensammlung des japanischen RIKEN Konsortiums wurde experimentell durch die Oligo-Capping 5´RACE Methode, auch 5´RLM-RACE genannt (RNA ligase-mediated rapid amplification of 5´ cDNA ends), generiert und sollte somit die vollständigen 5´UTR Bereiche aller amplifizierten Gen-transkripte beinhalten (Maruyama & Sugano, 1994; Suzuki & Sugano, 2003; Kodzius et al., 2004). Alternative Möglichkeiten zur genauen Bestimmung von 5´Transkriptenden bieten auch die Hochdurchsatzmethoden „CAGE“ („cap analysis gene expression“; Shiraki et al., 2003) und „5´ SAGE“ („5´-end serial analysis of gene expression“; Hashimoto et al., 2004).

Die TSS-Kartierungen wurden im nächsten Schritt als Startpunkt einer *in-silico* Analyse zur Identifizierung der assoziierten Core-Promotermotive, nach den Konsensussequenzregeln von Smale & Kadonaga (siehe Einleitung), verwendet.
Alle auf diese Weise identifizierten *DDX3Y* und *DDX3X* Transkriptklassen wurden anschließend auf ihr Expressionsprofil in verschiedenen Humangeweben analysiert. Dazu wurde eine Serie von RT-PCR Versuchen (Reverse Transcription Polymerase Chain Reaction) durchgeführt und analysiert, welche der ermittelten 5´UTR Transkriptvarianten einem spezifischen Core-Promoter zugeordnet werden können. Die kartierten TSS und alle weiteren Positionsangaben beziehen sich für die Gene *DDX3Y* und *DDX3X* auf den Y-BAC-Klon „475I1“[17], bzw. auf den X-BAC-Klon „RP13-13A3“[18]. Die mRNA Referenzsequenzen sind in der NCBI-Datenbank für *DDX3Y* unter GenBank Acc.No. NM_004660.3 und für *DDX3X* unter GenBank Acc.No. NM_001356.3 einsehbar. Eine Liste mit allen Positionsangaben übertragen auf die Human Y-Chromosom-[19] und X-Chromosom[20]-Referenzsequenz befindet sich im Anhang (Tab. A-2, A-3).

[17] *DDX3Y*: GenBank Acc.No. AC004474.1
[18] *DDX3X*: GenBank Acc.No. AL391647.16
[19] Y-Chr.: GenBank Acc.No. NC_000024.9 entsprechend der „GRCh37.1 Primary Reference Assembly“.
[20] X-Chr.: GenBank Acc.No. NC_000023.10 entsprechend der „GRCh37.1 Primary Reference Assembly“.

4.1.1. Kartierung der *DDX3Y* Transkriptionsstartstelle I (TSS-I)

Aus den bisherigen *DDX3Y* 5´RACE Experimenten ist bekannt, daß das 5´Ende der RACE Fragmente, sowohl in Hodengewebe als auch in Leukozyten, in Exon1 bei Position 55.042 im BAC 475I1 liegt (Hanstein, 2004). Diese 5´RACE Daten können somit als erste experimentelle Evidenz und Referenz für eine ubiquitäre Startstelle (TSS-I) der *DDX3Y* Transkription in Exon1, in allen Geweben, angesehen werden.
Um dies zu prüfen, wurden aus der DBTSS Datensammlung alle vorhandenen *DDX3Y* Transkripte extrahiert und in Tabelle 4-1 zusammengefasst. Parallel dazu wurden in Tabelle 4-1 auch die dazugehörigen Startstellen, entsprechend der Konsensussequenzen für das Initiator Core-Promotermotiv (YYANWYY) kartiert. Das im Initiator enthaltene, hochkonservierte Y-R Dinukleotid welches *in-silico* die TSS markiert, ist in Tabelle 1 unterstrichen.

Ein erstes Ergebnis dieser Auswertung ist die zu beobachtende Streuung der experimentell kartierten TSS-Positionen, über eine Region von 121 Nukleotiden, die sich 29 bis 149 Nukleotide stromaufwärts (5´) des *DDX3Y* ATGs im Exon1 befindet (TSS-I Region). Von den 27 unterschiedlichen TSS-Positionen aus der DBTSS konnten *in-silico* allerdings nur bei zweien eine perfekte Initiator-Konsensussequenz festgestellt werden (grün hervorgehoben in Tab. 4-1). 14 TSS-Positionen verfügen über das hochkonservierte „Y-R Dinukleotid“ (grau unterlegt in Tab. 4-1), welches als Mindestanforderung für ein funktionelles Initiatorelement gilt (Juven-Gershon et al., 2008; Juven-Gershon & Kadonaga, 2010). Elf der 27 Startstellen enthalten weder das Y-R Dinukleotid noch das vollständige Initiatormotiv. Es liegt somit die Vermutung nahe, daß der Sequenzbeginn der Einträge in der DBTSS Datenbank, trotz Einsatz der RLM-RACE Methode, nicht immer einer echten Transkriptionsstartstelle entspricht, oder daß es sich um alternative Initiationsmöglichkeiten handelt, die nicht durch die Initiator oder Y-R Regeln erfaßt werden.

Für die Größenberechnung der *DDX3Y* TSS-I Region werden TSS-Positionen ohne Y-R Dinukleotid oder Inr-Motiv, sowie solche aus Tumorgewebe nicht berücksichtigt. Die Initiationsregion wird dadurch auf 91 bp (BAC-Pos. 55.014 - 55.104) eingegrenzt. Diese Region stellt sich, anhand der Daten der DBTSS Datenbank, als die in vielen verschiedenen männlichen Geweben genutzte *DDX3Y* TSS-I Region dar. Die BAC-Position 55.042, welche als potentielle TSS in unserem 5´RACE Experiment erhalten wurde, stimmt mit einer Y-R Dinukleotid positiven DBTSS Position überein, die in vier verschiedenen Humangeweben (Hirn, Trachea, Milz, Thymus) als *DDX3Y* 5´Ende festgestellt wurde.

Tab. 4-1: Position potentieller Initiator Core-Promotermotive im Bereich des ubiquitären *DDX3Y* Transkriptstartes in Exon1 (TSS-I Region).

Die Positionsangaben in Spalte 1 entsprechen der +1 Position (= TSS) des Transkriptes, die wiederum dem dritten Nukleotid in einem Inr-Motiv entspricht. Die Positionsangaben beziehen sich auf den Human Y-BAC 475I1. Mismatches der ermittelten Sequenzen zur Inr-Konsensussequenz sind in Spalte 2 rot markiert. Das hochkonservierte Y-R-Dinukleotid ist in Spalte 2 unterstrichen und in Spalte 3 durch Pluszeichen gekennzeichnet. Initiatormotive mit 100% Übereinstimmung zur Konsensussequenz sind im Fettdruck und grün hervorgehoben. Weitere potentielle TSSs mit assoziiertem Y-R-Dinukleotid sind hellgrau dargestellt. TSSs die nicht der Inr-Seq. entsprechen bzw. kein Y-R Dinukleotid enthalten, sind farblich nicht unterlegt. In Spalte 4 wird angegeben, aus welchen Geweben/Zellen die Transkripte isoliert wurden. Die Zahlenangaben in Spalte 5 geben die Anzahl der Nennungen in der DBTSS an.

Position der 5´Enden der DBTSS Klone	Sequenz des pot. Inr-Elementes: YYANWYY	Y-R Regel für -1/+1 Position im Inr	Gewebe und Anzahl cDNA-Klone pro Gewebe	Gesamtzahl cDNA-Klone in DBTSS
54.997	ACGTTGT	+	Hauttumor	1
55.014	GTATAGC	+	Hirn	1
55.030	GGCCATA	-	Hirn, Gliom	2
55.031	GCCATAT	-	Darmkrebs	2
55.032	**CCATATT**	+	Hirn (2), Lunge, Trachea, Arterie, Gliom, NT2, Darmkrebs (2)	9
55.034	ATATTAC	+	Lunge, Arterie, Lebertumor	3
55.037	TTACCGC	+	Hirn	1
55.040	CCGCGTA	+	Immunzellen	1
55.042	GCGTAGG	+	Hirn (2), Trachea (2), Milz, Thymus	6
55.044	GTAGGCT	+	Thymus	1
55.053	CCAGTGC	+	Hirn (2), Thymus, Lebertumor, Gliom, unbestimmtes Gewebe	6
55.058	GCGCTTA	+	Hirn, Lungentumor	2
55.059	CGCTTAA	-	Lungentumor	2
55.062	TTAATAG	+	Thymus (2), Hoden	3
55.076	TCCAGTG	-	Hoden, Milz, Thymus, Trachea	4
55.077	CCAGTGT	+	Hirn (5), Trachea (5), Knorpel (4), Thymus (3), Zunge (3), Lunge (2), Milz, Blutzellen, Immunzellen, NT2 (2), Lebertumor	28
55.079	AGTGTAA	-	Knorpel	1
55.080	GTGTAAG	+	Hirn (3), Nebenniere, Knorpel, Haut	6
55.087	AGTTCCG	-	Thymus	1
55.088	GTTCCGC	-	Hirn	1
55.089	TTCCGCT	-	Trachea	1
55.090	TCCGCTA	-	Hoden	1
55.091	CCGCTAT	+	Hirn (4), Hoden (3), Thymus (3), Trachea (3), Arterie, Haut, Milz, Nierentumor, NT2	18
55.093	GCTATTC	-	Trachea, Blutzelle, NT2	3
55.094	CTATTCG	+	Hoden, Milz	2
55.104	**TCACACC**	+	Hirn	1
55.117	GGACTAC	-	Hirn	1

Sie liegt 10 Nukleotide stromabwärts (3´) zu der ersten perfekten Initiator Sequenz (CCATATT) in BAC-Position 55.032. Das zweite optimale Initiator-Motiv (TCACACC) liegt in BAC-Position 55.104 (Abb. 4-2). Zu den beiden perfekten Initiator-Elementen der *DDX3Y* TSS-I Region

konnten weder upstream noch downstream, weitere funktionelle Core-Promoter Elemente, wie TATA- oder DPE-Motiv, in den geforderten Abständen kartiert werden. Ein potentielles DPE der Sequenz GGTCCA, wurde im Abstand von +29 (BAC-Position 55.072), zu der *DDX3Y* TSS Position 55.044, identifiziert.

Für die Y-R positiven TSSs in den BAC-Positionen 55.077 und 55.091, für die die meisten Einträge in der DBTSS vorliegen (28 bzw. 18), konnte jeweils ein weiteres Core-Promoter-Motiv in entsprechender Distanz kartiert werden. Für die TSS in BAC-Position 55.077 wurde -31 Nukleotide stromaufwärts ein potentielles TATA-Box-Motiv der Sequenz „TTAATAGC“ identifiziert. Für die TSS in BAC-Position 55091 wurde ein potentielles DPE-Motiv der Sequenz „ACACCT“ in Position +28 gefunden (Abb. 4-2). Es sind also vier TSS-Position (BAC-Positionen 55.032, 55.077, 55.091, 55.104) in dieser Region erkennbar, die entweder durch ein optimales Inr-Motiv (55.032 & 55.104), oder durch Assoziation mit einem zweiten Core-Motiv (55.077 & 55.091) eine potentiell bevorzugte TSS-Nutzung erreichen können. Sie werden, von 5´ nach 3´, als *DDX3Y* TSS-Positionen **„*DDX3Y* TSS-Ia“** bis **„*DDX3Y* TSS-Id“** definiert (Abb. 4-2).

Da die meisten DBTSS Einträge in Tabelle 4-1 (101 von 108 Einträgen) zwischen den beiden perfekten Initiator-Motiven (BAC-Pos. 55.032 – 55.104) kartiert wurden, wird die gesamte 73 bp umfassende Region im Folgenden als **„*DDX3Y* TSS-I“** Region bezeichnet. Hieraus ergibt sich eine variable 5´UTR Sequenz, stromaufwärts des *DDX3Y* ATGs in Exon1, mit Längen zwischen 42 und 114 Nukleotiden.

Die Core-Promoter Analyse für die *DDX3Y* TSS-I Region zeigt, daß es sich hierbei, nach Definition von Sandelin et al. (2007) und Juven-Gershon et al. (2008), um den Typus eines breiten Core-Promoters handelt, in dem die Transkriptinitiation tatsächlich an mehreren Stellen, ohne deutlich dominante TSS-Position, möglich ist. Dieser breite Core-Promoter-Typ ist meist in (C,G)-reichen Sequenzregion anzutreffen. In Kapitel 4.6.2 wird gezeigt, daß im Bereich der *DDX3Y* TSS-I Region tatsächlich ein CpG-Island kartierbar ist.

Experimentell wurde das potentiell absolute 5´Ende der *DDX3Y* TSS-I Region über gezielte RT-PCR Analysen, entsprechend der Primerlokalisationen in Abb. 4-2, kartiert.
Die RT-PCR Analyse mit den Primerpaaren P1261/1286 sowie P2067/1286 zeigt keinen Unterschied zwischen cDNA aus Hodengewebe und cDNA Proben aus männlichen Leukozyten (Abb. 4-1). Diese Sequenzregion ermöglicht also, wie aus Tab. 4-1 erwartet, eine *DDX3Y* Transkription in beiden Geweben bzw. Zellen. Mit dem Primerpaar P2405/1286 (Position 54.925), konnten nur im Hodengewebe spezifische *DDX3Y* PCR-Amplifikate gefunden werden. Mit dem 40 bp stromaufwärts von P2405 gelegenen Forward-Primer P2389 (Position 54.885)

und dem gleichen Reverse-Primer P1286 konnten auch in cDNA Proben aus Hodengewebe keine *DDX3Y* Transkripte mehr festgestellt werden (Abb. 4-1).

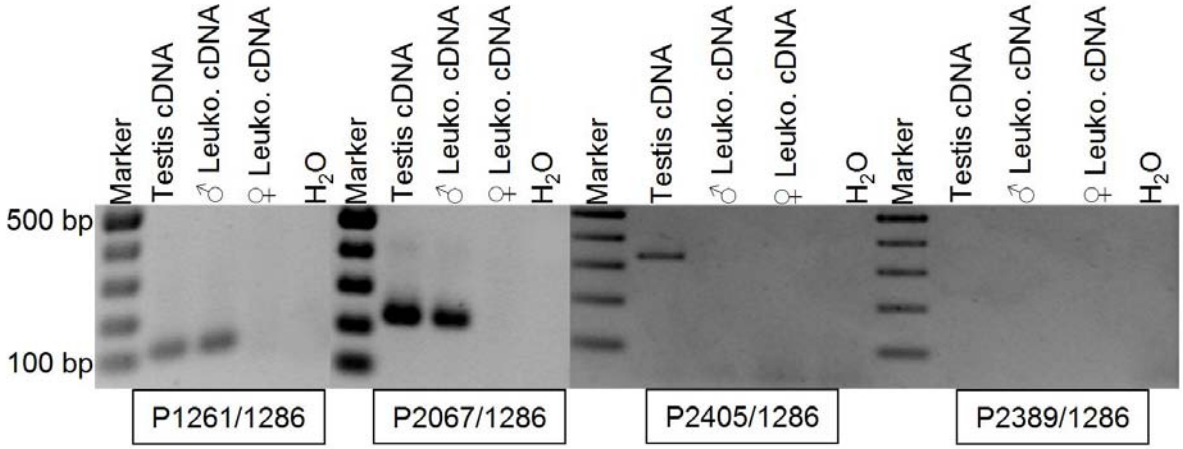


Abb. 4-1: RT-PCR Analyse zur Kartierung der absoluten Grenze der *DDX3Y* TSS-I Region.
Links das Ergebnis mit Primerpaar P1261/1286 (PCR-Produkt = 102 bp), Mitte links mit P2067/1286 (PCR-Produkt = 222 bp), Mitte rechts mit P2405/1286 (erwartetes PCR-Produkt = 328 bp) und rechts mit P2389/1286 (PCR-Produkt errechnet = 368 bp). Zum Ausschluß von Kreuzreaktionen mit Transkripten des gametologen *DDX3X* Gens wurden die Primerpaare hier ebenso mit der cDNA weiblicher Leukozyten analysiert. 35 PCR Zyklen. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

Die entsprechende Startstelle sollte in Hodengewebe also zwischen den Primerpositionen P2389 und P2405 zu finden sein. Diese scheinbar hodenspezifische TSS-I Region verlängert die „*DDX3Y* Exon1-Sequenz“ demnach im Hodengewebe um etwa 100 Nukleotide.

Zwei kanonische Core-Promoter Motive konnten in diese hodenspezifisch verlängerte TSS-I Region kartiert werden. BAC-Position 54.911 wurde als potentielles Initiator-Motiv mit einem Mismatch (CTAAACG) identifiziert (Abb. 4-1). In Position +29, relativ zu dieser Initiator-Sequenz, wurde ein potentielles DPE, mit zwei Mismatches (GGTAGC) zur Konsensussequenz, lokalisiert (Abb. 4-2). Dieses Core-Promoter Modul aus zwei Motiven scheint also hauptsächlich im Hodengewebe funktionell zu sein. Lediglich schwache Amplifikate wurden ebenfalls in anderen Humangeweben gezeigt (Zimmer J., unveröffentlichte Resultate). Diese TSS wird daher als **„*DDX3Y* TSS-I-($T_{(estis)}$)“** definiert.

Für *DDX3Y* kann somit zusammenfassend gesagt werden, daß eine ubiquitär genutzte TSS-I Region mit 5´UTR Längen zwischen 42 (BAC-Position 55.104) und 114 (BAC-Position 55.032) Nukleotiden, sowie eine vornehmlich in Hodengewebe genutzte distale TSS-I-(T) Region (BAC-Position 54.911), vorhanden ist. Die um 121 Nukleotide längere 5´UTR (235 Nukleotide) dieser TSS-I-(T) *DDX3Y* Transkriptklasse weist als Besonderheit zwischen TSS-I-(T) und TSS-Ia eine auffallend G-reiche Region mit einem nicht-kodierenden $(TGG)_4$ Trinukleotidrepeat (TNR) auf (Abb. 4-2, siehe hierzu Kapitel 4.11.2).

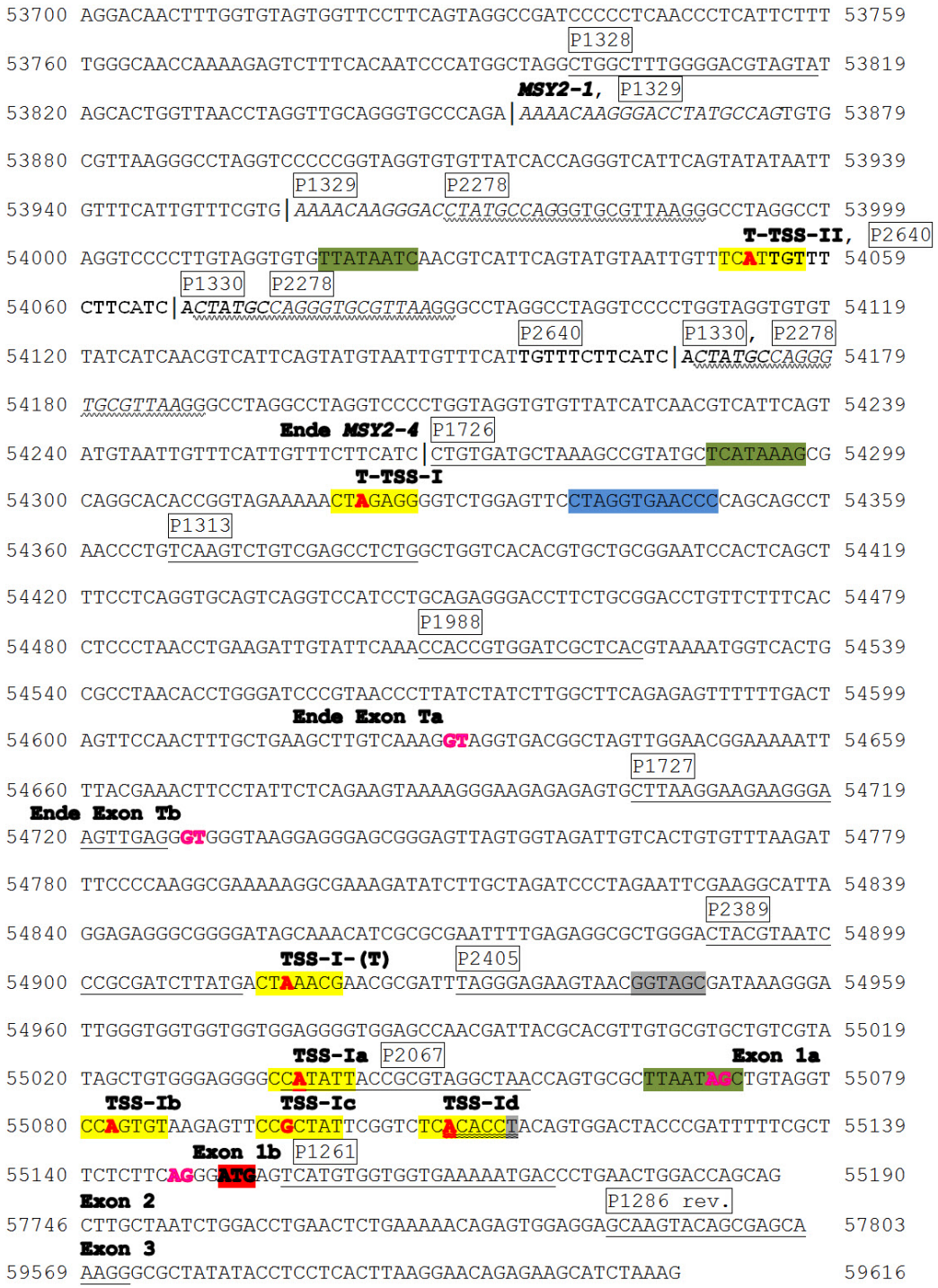


Abb. 4-2: Genomische Sequenz des *DDX3Y* Genlocus von Position 53.700 bis 55.190 (Ende Exon1) im BAC 475I1.

Die für die RT-PCR-Analyse der *DDX3Y* TSS-Regionen verwendeten Primer sind in der Sequenz unterstrichen. Primer P2278 ist zur besseren Unterscheidung zu P1330 gewellt unterstrichen, Primer P2640 ist im Fettdruck dargestellt. Die Primer P1329 und P1330 sind kursiv dargestellt. Der Reverseprimer P1286 ist überlappend in Exon2/3 angegeben. Die identifizierten Core-Promoter Motive sind alle farblich hervorgehoben. Initiator-Motive sind gelb unterlegt und die TSSs innerhalb der Inr-Motive sind rot markiert. Potentielle TATA-Box-Motive sind grün, MTE-Motive sind blau und DPE-Motive sind grau unterlegt. Das TSS-Id Inr-Motiv in 55.104 fällt mit einem DPE-Motiv zusammen. Das DPE ist zusätzlich doppelt gewellt unterstrichen. Das *DDX3Y* Translationsstartcodon ATG ist rot hervorgehoben. Die Splicesignale innerhalb von Exon-T bzw. Exon1 sind in pink und Fettdruck markiert. Die vier *MSY2* Repeateinheiten sind durch senkrechte Striche getrennt. Die Primerbezeichnungen beziehen sich auf die interne Labornomenklatur und sind mit einem Kasten dargestellt.

4.1.2. Kartierung der potentiellen *DDX3Y* TSS-Region im *MSY2* Minisatelliten-repeat

Erste Hinweise für eine weitere hodenspezifische *DDX3Y* Transkriptklasse mit deutlicher 5´ Verlängerung lieferten die *DDX3Y* 5´RACE Produkte, die in der Diplomarbeit von Hanstein (2004) ausschließlich hodenspezifisch isoliert werden konnten. Per Sequenzanalyse konnten Abschnitte des Y-Chromosom-spezifischen Minisatellitenrepeats *MSY2* (*DYS440*; Bao et al., 2000), am 5´Ende dieser neuen Transkriptklasse identifiziert werden. Auf Grund der entfernten Lage, *MSY2* befindet sich zwischen 1292 und 884 bp stromaufwärts des *DDX3Y* ATGs in Exon1, legten diese Daten eine zweite, distale *DDX3Y* Promoter-Region nahe, die nur in Hodengewebe geöffnet und genutzt werden kann.
Aus den Sequenzierungen der PCR-Produkte wurde ebenfalls erkannt, daß die 5´ bis *MSY2* verlängerten Hoden-Transkripte alle in einem spezifischen Muster gesplict werden und ohne Veränderung des Leserahmens, mit *DDX3Y* Exon1 verknüpft werden (Hanstein, 2004; Rauschendorf et al., 2011).

Für einen möglichen *DDX3Y* Transkriptstart in *MSY2* sind bisher keine Einträge in den NCBI-, Ensembl- oder DBTSS-Datenbanken vorhanden.
Der potentielle *DDX3Y* Transkriptstart in *MSY2* wird daher zunächst auf Anwesenheit potentieller Core-Promotermotive, mit Fokus auf TATA-Box- und Initiator-Motive, hin untersucht.

Aus dem ClustalW Sequenzalignment der vier *MSY2* Repeateinheiten ist deutlich zu erkennen, daß potentielle Initiator-Elemente in allen Repeateinheiten hoch repetitiv vorliegen (Abb. 4-3). Es sind 28 potentielle Initiatormotive zu unterscheiden, die in acht Blöcke unterteilt werden können (Tab. 4-2). In keinem der Blöcke ist eine perfekte Übereinstimmung zur kompletten Konsensussequenz „YYANWYY“ zu finden ist. Die Inr-Sequenz lässt sich für alle Blöcke zu der Formel **„YYABNNH“** (B = G/C/T; H = A/C/T) vereinfachen und ergibt so ein wiederkehrendes Muster innerhalb der Repeateinheiten (Abb. 4-3). In Repeat 1 sind, auf Grund von Sequenzunterschieden zu den Repeateinheiten 2, 3 und 4, keine Inr-Motive in den Blöcken II, VII und VIII vorhanden.

Durch die Deletion der ersten elf Nukleotide in Repeat 3 und 4 entstehen an der Grenze von Repeateinheit 2→3 und 3→4 zwei neue, potentielle Initiator-Motive (Block VIII) der Sequenz TCACTAT mit einem Mismatch (rot markiert), die in Repeat 1 und 2 nicht vorhanden sind (Abb. 4-3). Am 3´Ende von Repeat 4 ist ein weiterer Initiator TCATCCT möglich (Nr. 29; 1 Mismatch), der durch die zwei folgenden Nukleotide (unterstrichen) stromabwärts des Repeats komplettiert wird.

Die Initiator-Motive in Block IV mit der Sequenz TTATC/AAC/T (zwei Mismatches) können ebenso eine potentielle TATA-Box Sequenz (TTATC/AAC/TC) darstellen. In Repeateinheit 2 wird durch Transversion von C zu A an Position 5 dieses TATA-Box Motives die Sequenz „TTATAATC“ erzeugt. Dieses TATA-Box Motiv entspricht, im Vergleich zu den anderen TATA-Motiven in Block IV, eher den Core-Promoterregeln von Smale & Kadonaga (2003), so daß für diese Position eher eine TATA-Box und kein Inr-Motiv zu definieren ist.

Das erste T dieser TATA-Box (Block IV) ist in Position -34 relativ zum +1 A-Nukleotid des Inr-Motives in Block VI lokalisiert. Es befindet sich damit in der geforderten Distanz für kooperierende TATA- und Inr-Elemente in einem Core-Promoter.
Ein zweiter Block potentieller TATA-Box Motive der Sequenz TATA/GTAAT befindet sich zwischen den Inr-Motiv Blöcken V und VI. An Position vier des TATA-Motives, in Repeateinheit 1, ist im Gegensatz zu den Repeateinheiten 2, 3 und 4, ein optimales Adenin vorzufinden.
Nur in den Repeateinheiten 2 und 3 befindet sich das erste T-Nukleotid dieser TATA-Motive im optimalen Abstand der Initiator-Motive (Position -28 relativ zum +1 Adenin) in Block VIII. Diese Kombination aus TATA-Box und den beiden Inr-Motiven in Block VIII, die die Repeatgrenzen überlappen, ist nur in den Repeats 3 und 4, auf Grund der Deletion von elf Nukleotiden an der 5´ Seite der Repeateinheiten, möglich. Die Entstehung dieser beiden Core-Module aus zwei Motiven an den Repeatgrenzen ist humanspezifisch, da Bao et al. (2000) bereits für einige andere Primaten zeigen konnten, daß diese nur über zwei *MSY2* Repeateinheiten verfügen.

Das TATA-Box Motiv in Repeateinheit 1 (zwischen Block V und VI) mit der Sequenz „TATATAAT“ entspricht, im Unterschied zu den entsprechenden TATA-Box Motiven in den Repeats 2 bis 4, in allen Positionen der Konsensussequenz. Ein minimales Inr-Element mit dem YR-Dinukleotid „CA“ in Position 53.960 ist im 5´ Bereich von Repeat 2 vorhanden (Abb. 4-3). Diese TATA-Box erfüllt ebenfalls die korrekte Distanz mit Lokalisation an Position -30 relativ zum CA-Dinukleotid.

Als drittes wichtiges Core-Promoterelement konnte ein potentielles DPE-Motiv der Sequenz „GGTCCC“ identifiziert werden, welches der DPE-Konsensussequenz „DSWYVY“ entspricht (Abb. 4-3). Da das DPE nicht in der Lage ist alleine den PIC zu rekrutieren, ist die Anwesenheit eines Initiators, oder einer TATA-Box in entsprechender Entfernung notwendig. Initiator-Motive in entsprechender upstream Entfernung zu dem „GGTCCC“- Motiv sind näherungsweise nur für Block I zu finden. Für Repeat 1 beträgt die Entfernung +24, für Repeat 2 +30 und für Repeat 3 und 4 befindet sich das DPE an Position +33. Somit ist für das potentielle DPE-Motiv festzuhalten, daß keines der Initiatormotive in optimaler Entfernung lokalisiert ist.

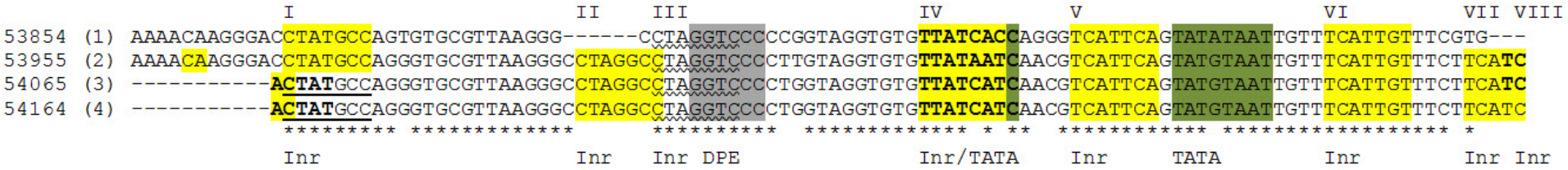


Abb. 4-3: Core-Promoter Motive in der *MSY2*-Repeatsequenz stromaufwärts des *DDX3Y* Gens.

Die acht Blöcke repetitiver Initiator-Elemente sind mit römischen Ziffern versehen. Initiator-Motive sind gelb dargestellt. Die Initiatoren an den Grenzen der Einheiten 2/3 und 3/4 sind in Fettschrift hervorgehoben (Block VIII). Die Initiator-Motive aus Block I sind in Repeat 3 und 4 unterstrichen. Block III ist in allen Repeateinheiten gewellt unterstrichen dargestellt. Potentielle TATA-Sequenzen sind grün bzw. zusätzlich in Fettdruck (in Inr Block IV) markiert. DPEs sind grau unterlegt. Zur deutlicheren Darstellung des homologen DPE-Motives in den vier Repeateinheiten wurde das Palindrom „CCTAGG“ in Repeat 1 mit dem verdoppelten Palindrom in den Repeats 2 bis 4 übereinandergelegt. Am Ende von Repeat 4 ist ein Initiator „TCATCCT“ mit den zwei darauf folgenden Nukleotiden außerhalb des Repeats möglich. Unterhalb der Blöcke ist die Art des Core-Promoter Motives angegeben.

Inr-Block	***MSY2* Repeateinheit**
I	1 bis 4
II	2 bis 4
III	1 bis 4
IV	1 bis 4
V	1 bis 4
VI	1 bis 4
VII	2 bis 4
VIII	2→3, 3→4
gesamt	**28**

Tab. 4-2: Übersicht der Inr-Motive in den vier *MSY2* Repeateinheiten.

Angegeben wird die Lage des jeweiligen Inr-Blockes in Bezug zu den vier Repeateinheiten.

Module aus Inr- und DPE-Motiv kommen somit nicht in Frage, da der Abstand zwischen Inr- und DPE-Motiv funktionell kritisch und, im Gegensatz zum Abstand zwischen TATA- und Inr-Motiv (von -28 bis -34), nicht variabel ist (Juven-Gershon et al., 2008). Eine Kombination aus TATA- und DPE-Motiv ist dagegen möglich. Die TATA-Motive zwischen Block V und VI in Repeat 2 und 3 befinden sich in der geforderten Distanz zu den DPE-Motiven in Repeateinheit 3 und 4. Die TATA-Motive liegen jeweils 61 Nukleotide stromaufwärts der DPE-Motive. Aus der Addition der Abstände von TATA zu Inr und Inr zu DPE ergibt sich ein Fenster von 56 bis 62 Nukleotiden für den Abstand von TATA- zu DPE-Motiv.

In der *MSY2* Repeat-Region (BAC-Position 53.854 - 54.262) konnten somit 28 potentielle Initiator-Motive, acht TATA-Box Motive und vier DPE-Motive kartiert werden (Abb. 4-3). Welches dieser Motive bzw. welche Motivkombination potentiell für die hodenspezifische *DDX3Y* Transkription mit Start im *MSY2*-Repeat verantwortlich ist, soll im folgenden Abschnitt durch gezielte RT-PCR Experimente analysiert werden.

Da bekannt ist, daß die Klonierung repetitiver DNA zu Instabilität und Rearrangements der Sequenzen führen kann (Neil et al., 1990; Glover & Hames, 1995), wurde zur Analyse der potentiellen *MSY2* TSS-Region ein RT-PCR Ansatz mit spezifischen Oligonukleotiden (Abb. 4-5), einem 5´RACE Experiment vorgezogen.

Aus Abb. 4-4 ist zu erkennen, daß mit den Primerkombinationen P1328/1286 und P1329/1286 weder in cDNA aus Hodengewebe noch in cDNA männlicher Leukozyten RT-PCR Amplifikate beobachtet werden können. Auf Grund dessen kann ein *DDX3Y* Transkriptstart stromaufwärts von *MSY2* und im 5´ Bereich der Repeateinheit 1 ausgeschlossen werden. Mit den Primerpaaren P1330/1286 (Repeat 3 & 4) und P2278/1286 (Repeat 2, 3 & 4), die am 5´ Ende der Repeateinheiten binden (siehe Abb. 4-5), konnten jeweils zwei hodenspezifische PCR-Produkte identifiziert werden. D.h., der 5´ Bereich des *MSY2* Repeats 3 ist bereits Bestandteil des hodenspezifischen *DDX3Y* Transkriptes. PCR-Produkte mit einer Amplifikatgröße von über 800 bp sind nicht zu erkennen, was der Fall für eine P2278 Primerbindung in Repeat 2 wäre. Eine Initiation im 5´ Bereich von Repeateinheit Nr. 2 ist daher unwahrscheinlich.

Die potentielle *DDX3Y* Transkriptionsstartstelle in *MSY2* sollte daher zwischen den Primerpositionen von P2278 in Repeat 2 und P1330 in Repeat 3 zu suchen sein. Um zu klären, ob die Transkriptinitiation zu Beginn von Repeateinheit 3 oder im 3´ Bereich von Repeat 2 stattfindet, wurde Primer P2640 entwickelt, der spezifisch Repeat 2 und 3 im 3´ Bereich markiert (Abb. 4-5).

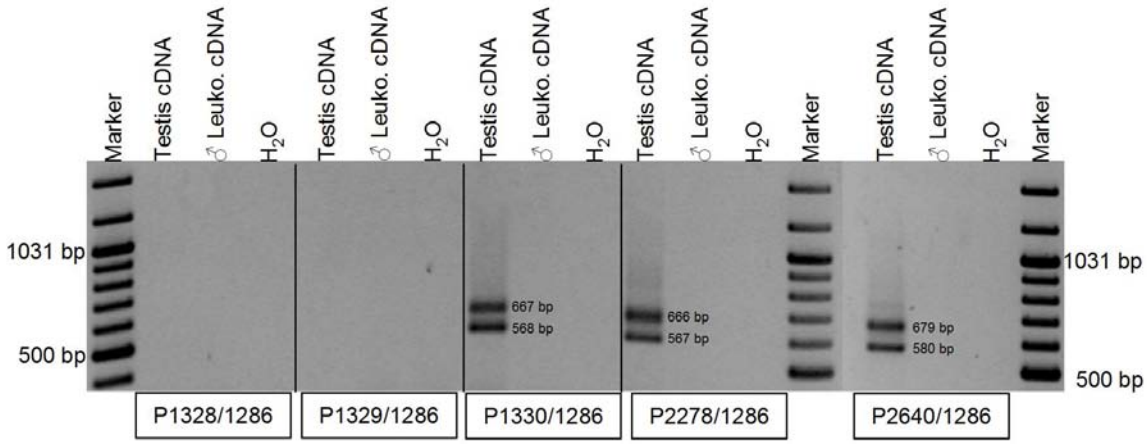


Abb. 4-4: RT-PCR Experimente zur Eingrenzung der *DDX3Y* Transkriptionsstartstelle im *MSY2* Repeat.

Lokalisation der verwendeten Primer siehe Abb. 4-2 u. 4-5. RT-PCR-Produktgrößen sind in Tab. 4-3 aufgelistet. 35 PCR-Zyklen. Verwendeter DNA-Marker ist eine 100bp DNA-Leiter.

In Abb. 4-4 sind zwei deutliche RT-PCR Amplifikate mit der Primerkombination P2640/1286 sichtbar, was auf einen *DDX3Y* Transkriptionsstart im 3´ Bereich von *MSY2* Repeat 2 hindeutet.

Da die Primer P2278 und P2640 an je zwei Positionen im *DDX3Y* Transkript binden können und drei Exon-T Splicevarianten (Ta-1b, Ta-1a, Tb-1b; siehe Rauschendorf et al., 2011) existieren, ergeben sich je Primer sechs unterschiedliche Transkripte (siehe Tab. 4-3), die in einer RT-PCR Reaktion amplifiziert werden und sich in ihrer PCR-Amplifikatgröße auf dem Agarosegel überlagern können.
Um festzustellen, ob es sich bei dem RT-PCR Amplifikat (P2640/1286) mit einer ungefähren Größe von 680 bp (Abb. 4-4), um verschiedene Exon-T Splicevarianten oder um Transkripte mit unterschiedlichem Startpunkt handelt, wurde für P2640 und für P2278 eine zweite RT-PCR Serie mit dem Reverseprimer P1617 durchgeführt. Dieser ist innerhalb von Exon-T lokalisiert, so daß *DDX3Y* Exon-T Splicevarianten vermieden werden und Transkripte mit unterschiedlichem Startpunkt getrennt dargestellt werden können.

Das RT-PCR Experiment mit Primerpaar P2640/1617 zeigt zwei deutliche Amplifikate der berechneten Größe von 340 bp und 439 bp (Abb. 4-6).

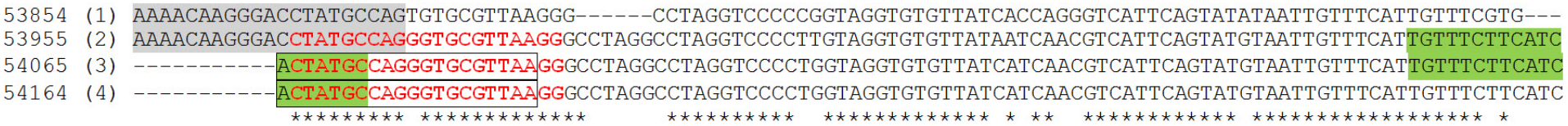


Abb. 4-5: *MSY2* Repeat-Sequenz (Position 53.854 – 54.262 in BAC 475l1) mit eingezeichneten Primersequenzen zur RT-PCR Analyse.
Primer P1329 ist hellgrau unterlegt und spezifisch für die *MSY2* Repeateinheiten 1 und 2. P1330 ist durch einen Kasten markiert und spezifisch für die Repeateinheiten 3 und 4 entwickelt. P2278 ist rot im Fettdruck hervorgehoben und markiert die Repeateinheiten 2, 3 und 4 jeweils im 5´Bereich. P2640 ist grün dargestellt und so entwickelt, daß nur die Repeats 2 und 3 erkannt werden (siehe auch Tab. 4-3).

Primer Nr.	Primerposition	*MSY*-1	*MSY*-2	*MSY*-3	*MSY*-4
P1328	53.798	---	---	---	---
P1329	53.854, 53.955	X	X	---	---
P1330	54.065, 54.164	---	---	**X** Ta-1b: 667 bp Ta-1a: 744 bp Tb-1b: 766 bp	**X** Ta-1b: 568 bp Ta-1a: 645 bp Tb-1b: 667 bp
P2278	53.967, 54.066, 54.165	---	X [Ta-1b: 765 bp] [Ta-1a: 842 bp] [Tb-1b: 864 bp]	**X** Ta-1b: 666 bp Ta-1a: 743 bp Tb-1b: 765 bp	**X** Ta-1b: 567 bp Ta-1a: 617 bp Tb-1b: 666 bp
P2640	54.053, 54.152	---	**X** Ta-1b: 679 bp Ta-1a: 756 bp Tb-1b: 778 bp	**X** Ta-1b: 580 bp Ta-1a: 657 bp Tb-1b: 679 bp	---

Tab. 4-3: Zusammenfassung der *MSY2* RT-PCR Experimente und Angabe der PCR-Amplifikatgrößen.
Antisenseprimer ist für diese RT-PCR Serie immer P1286. Primer P1328 ist vor dem *MSY2* Repeat lokalisiert und dient als Kontrolle, ob Initiation vor dem *MSY2* Repeat möglich ist. X markiert die Komplementarität des Primers an entsprechende Repeateinheit, keine Primerbindung ist durch Striche angedeutet. Die grün unterlegten Felder geben an, daß mit diesen Primern spezifische RT-PCR Amplifikate in Human-Hodengewebe identifiziert werden konnten. PCR-Produktgrößen für P2278 in *MSY2-2* sind in eckigen Klammern angegeben, da keines dieser möglichen Produkte detektiert werden konnte.
In Abb. 4-4 fällt auf, daß mit den gewählten Primerkombinationen stets nur zwei deutlich distinkte PCR-Banden, trotz jeweils zwei möglicher Primerbindungsstellen und drei unterschiedlicher Exon-T Splicemöglichkeiten, darstellbar sind. Die hier berechneten PCR-Amplifikatgrößen machen deutlich, daß die Amplifikate der Minor-Spliceprodukte Ta-1a und Tb-1b durch die Haupt-Splicevariante Ta-1b überlagert werden und so nur zwei anstelle von sechs PCR-Banden im Agarosegel sichtbar werden.

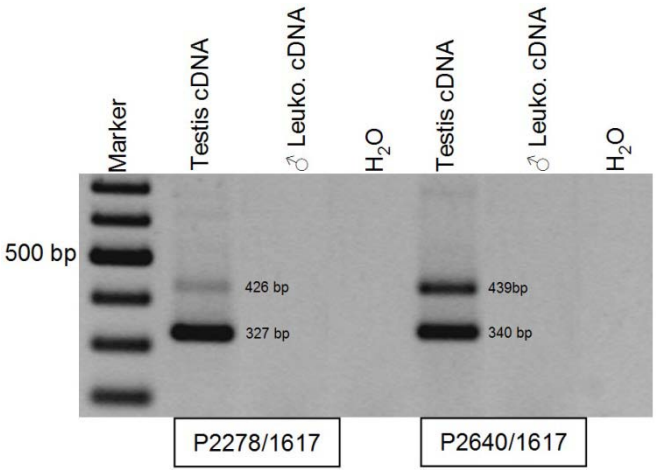


Abb. 4-6: RT-PCR Analyse zur Bestimmung der *DDX3Y* Transkriptionsinitiation in *MSY2*. Berechnete RT-PCR Produkt-größen für P2278/1617 = 327 und 426 bp. Amplifikat-größen für P2640/1617 = 340 und 439 bp. 35 PCR Zyklen. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

D.h. für Primer P2640 sind beide möglichen Primer-Bindungsstellen in den *DDX3Y* Transkripten vorhanden. Die Initiation erfolgt demnach bereits in *MSY2* Repeat 2, stromaufwärts der ersten P2640 Primerlokalisation. Die Analyse mit der Primerkombination P2278/1617 bestätigt dieses Ergebnis, da hier ebenfalls zwei Amplifikate zu detektieren sind, auch wenn das größere Amplifikat von 426 bp nur schwächer dargestellt werden konnte. Die dritte mögliche P2278 Primer-Bindungsstelle, zu Beginn von *MSY2* Repeat 2, kann in diesem RT-PCR Experiment, wie schon in Abb. 4-4, nicht dargestellt werden. (Abb. 4-6). Durch die Kombination von *in-silico* Core-Promoter-Motiv Kartierung und repeatspezifischen RT-PCR Analysen konnte die distale, hodenspezifische *DDX3Y* Initiationsregion somit auf den 3´ Bereich des zweiten, bzw. auf den 5´ Bereich des dritten *MSY2* Repeats eingegrenzt werden.

Das potentielle Core-Promoter Modul für diese Transkriptionsinitiation besteht somit aus der TATA-Box in Block IV und dem Inr-Motiv in Block VI. Die Tatsache, daß die TATA-Box Sequenz im *MSY2* Repeat 2, Block IV, den Konsensussequenz-Regeln entspricht, macht dieses Modul zu einem optimalen Kandidaten für die gesuchte *DDX3Y* TSS innerhalb der *MSY2* Minisatelliten -region. Die daraus resultierende hodenspezifische TSS wird in Position 54051 im BAC 475I1 kartiert (Abb. 4-2).

4.1.3. Kartierung weiterer potentieller TSS-Regionen im *DDX3Y* Exon-T

In der DBTSS Datensammlung konnten keine hodenspezifischen *DDX3Y* Transkripte mit *MSY2* Anteil identifiziert werden. Es konnten aber 5 Transkripte mit *DDX3Y* Exon-T Sequenz fest-gestellt werden, die ihren Startpunkt proximal zu *MSY2* haben. Diese als hodenspezifisch klassifizierten *DDX3Y* cDNA-Klone hatten in BAC-Position 54.316 (2 Einträge) bzw. in Position 54.317 (3 Einträge) ihre experimentell definierte Startposition. In Position 54.316 konnte mit der Sequenz „ACTAGAG“ kein Hinweis auf ein funktionelles Inr-Motiv gefunden werden. Weder die

Konsensussequenz (fünf Mismatches, rot markiert), noch die Y-R-Dinukleotidregel läßt auf ein potentielles Initiatormotiv schließen.
Für Position 54.317 konnte ein potentielles Initiatormotiv, der Sequenz „CTAGAGG“ (zwei Mismatches, rot markiert), kartiert werden, welches die Y-R-Regel mit dem Dinukleotid TA erfüllt. An Position -32, relativ zum Initiator-Motiv mit TSS an Position 54.317, ist ein potentielles TATA-Box Motiv (TCATAAAG) mit einem Mismatch an Position 2 (rot markiert) zu finden. Durch die in der Einleitung (Kapitel 1.6) beschriebene Konkurrenz von TATA-Box und Initiator, um die Positionierung des PIC, könnte das Vorhandensein von Transkripten mit 5´Beginn an Position 54.316 damit erklärbar sein, obwohl diese TSS kein Y-R Dinukleotid aufweist.

Eine Datenbankrecherche nach vergleichbaren *DDX3Y* Transkripten in der NCBI-Datenbank ergab einen Eintrag für ein Transkript mit Start an Position 54.317 im BAC 475I1 (Testis-cDNA Klon BC034942). Dieser cDNA-Klon repräsentiert ein *DDX3Y* Transkript mit dem bekannten Exon-Ta-1b Splicingmuster.

Auf der Suche nach weiteren potentiellen Core-Promoter Elementen, konnte ein potentielles MTE an Position +18 relativ zur TSS in Pos. 54317 festgestellt werden. In der hier vorgeschlagenen MTE-Sequenz „CTAGGTGAACCC“ sind fünf Abweichungen zur Mammalia MTE-Konsensussequenz „SVAGCSSRGCGS“ zu finden (Jin et al., 2006). Die Abweichungen zur Konsensussequenz können eine Reduktion der Transkriptionsaktivität bzw. der Bindungsfähigkeit zum TFIID-Komplex zur Folge haben. Die relative Position zum +1 Nukleotid im Initiator-Motiv ist mit +18 exakt gewahrt, was große funktionelle Bedeutung für die synergistische Wirkung des MTE-Motives hat (Lim et al., 2004).

Zusammenfassend läßt sich somit proximal zu *MSY2*, ein zweites hodenspezifisches Core-Promotermodul mit TSS in BAC-Position 54.317, bestehend aus TATA-Inr-MTE, vorgeschlagen. Es ist in der genomischen *DDX3Y* Sequenz in Abb. 4-2 dargestellt (dunkelgrün-gelb-hellblau). Exon-T hat damit zwei hodenspezifische TSS-Positionen. Sie werden als **„*DDX3Y* T-TSS-I“** (BAC-Pos. 54.317) und **„*DDX3Y* T-TSS-II“** (BAC-Pos. 54.051) definiert.

Wie in Abb. 4-7 zu sehen ist, ergab das RT-PCR Experiment mit den drei Forward-Primer P1726 (BAC-Pos. 54.263), P1313 (BAC-Pos. 54.362) und P1988 (BAC-Pos. 54.502) ausschließlich RT-PCR Amplifikate in Hodengewebe. Das hodenspezifische Muster aus drei Banden stellt die drei Splicevarianten (Ta-1b, Ta-1a, Tb-1b) der *DDX3Y* Transkripte mit Start in Exon-T dar.
Das positive RT-PCR Ergebnis mit dem stromaufwärts von T-TSS-I gelegenen Primer P1726 bestätigt nochmals die Ergebnisse der Sequenzanalyse, daß Transkripte, die im *MSY2*

Minisatelliten Repeat beginnen, durchgehend bis zum Ende von Exon-T sind. Die T-TSS-I befindet sich ~ 830 Nukleotide stromaufwärts des *DDX3Y* ATGs.

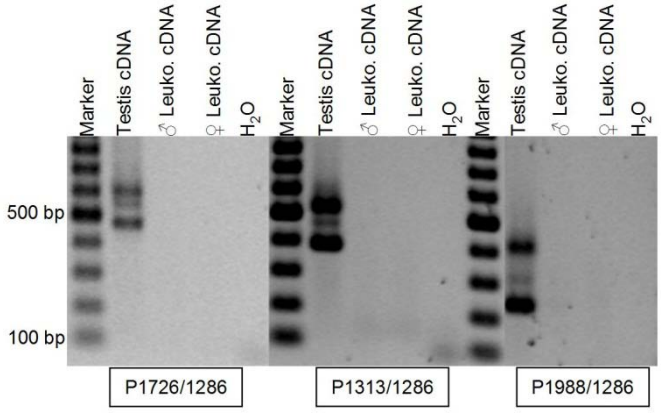


Abb. 4-7: RT-PCR Analysen zur Transkript-initiation in der *DDX3Y* Exon-T Region.
Links das Ergebnis mit Primerpaar 1726/1286 (PCR-Produkte: Ta-1b = 470bp; Ta-1a = 547bp; Tb-1b = 569bp), mittig mit 1313/1286 (PCR-Produkte = Ta-1b = 371bp; Ta-1a = 448bp; Tb-1b = 470bp) und rechts mit 1988/1286 (PCR-Produkte = Ta-1b = 231bp; Ta-1a = 308bp; Tb-1b = 330bp). Zum Ausschluß von Kreuzreaktionen mit Transkripten des gametologen *DDX3X* Gens, wurden die Primerpaare auch mit cDNA aus weiblichen Leukozyten getestet. 35 PCR Zyklen. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

Je nach Splicevariante haben die Transkripte mit Exon-T eine 5´UTR Länge von ~ 310 bis 410 Nukleotiden. Die abschließenden Ergebnisse für die Kartierung der *DDX3Y* Transkriptions-startstellen, die aus der Kombination der Core-Promoteranalyse und den RT-PCR Experimenten erhalten wurden, sind in den Abbildungen 4-2 und 4-8 zusammengefasst. Es konnten vier Regionen (TSS-I, TSS-I-(T), T-TSS-I, T-TSS-II) beschrieben werden, in denen die Initiation von *DDX3Y* Transkripten erfolgen kann, wobei zwei Regionen (T-TSS-I, T-TSS-II) ausschließlich in Hodengewebe genutzt werden. Für alle vier Bereiche konnte eine gute Übereinstimmung von experimentell gewonnenen RT-PCR Ergebnissen sowie *in-silico* erhobene Daten festgestellt werden. Somit kann von mindestens zwei unterschiedlichen *DDX3Y* Promoter-Domänen ausgegangen werden, die distale Region mit hodenspezifischem und die proximale Region mit ubiquitärem Transkriptionsstart.

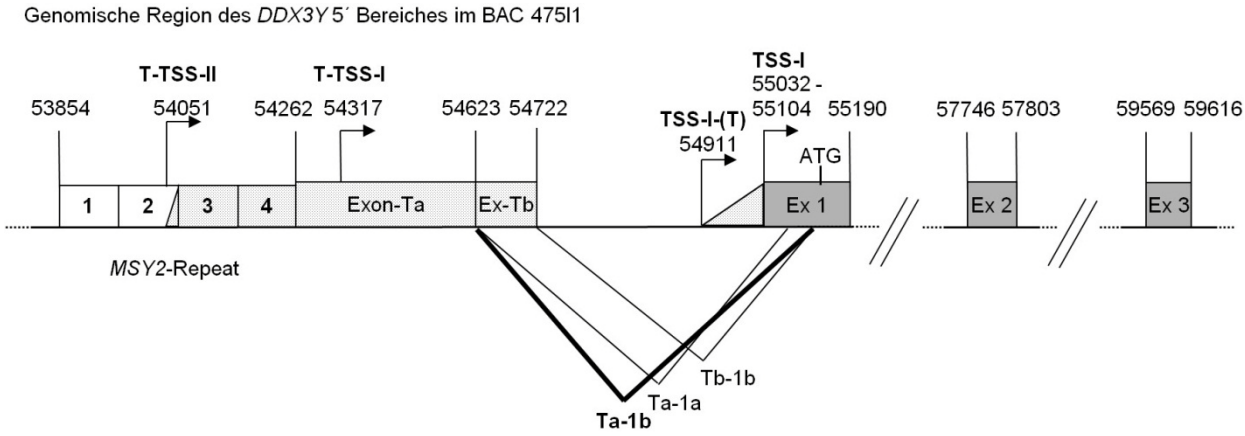


Abb. 4-8: Graphische Zusammenfassung der *DDX3Y* Transkriptionsinitiationsstellen.
Eingezeichnet sind die vier kartierten *DDX3Y* TSS Regionen mit jeweiliger Position im BAC 475I1, sowie das Splicingmuster der 5´ verlängerten Transkripte mit Exon-T. Die 5´UTR Bereiche, die hodenspezifisch bzw. überwiegend im Hoden exprimiert werden, sind gepunktet eingezeichnet. Die ubiquitär exprimierten Exone sind grau ausgefüllt dargestellt.

4.2. Keimbahnkomplexität der *DDX3X* Transkripte und Kartierung der dazugehörigen Core-Promotermotive

Analog zur *DDX3Y* Analyse wurde auch für den 5´ Genbereich von *DDX3X* eine Kartierung der potentiellen TSS-Regionen, unter Verwendung von Datenbankrecherche, *in-silico* Core-Promoteranalyse und anschließender experimenteller Kontrolle durch geeignete RT-PCR Experimente, durchgeführt.
Zur ersten Analyse auf potentielle Gewebespezifitäten von *DDX3X* Transkripten dienten cDNA Proben aus Hodengewebe und männlichen, sowie weiblichen Leukozyten. Die weiblichen Leukozyten wurden in die Analyse aufgenommen, um auch eventuelle Unterschiede in der TSS-Nutzung zwischen männlichen und weiblichen Leukozyten zu erkennen.
Als Einstiegsexperiment in die *DDX3X* TSS Analyse dienten bekannte 5´RACE Produkte aus dem Forschungspraktikum von Fornoff (unveröffentlichte Resultate). So konnte für männliche Leukozyten ein 5´RACE cDNA-Klon mit 5´Ende an Position 30.642 und für Hodengewebe an Position 30.637 im BAC RP13-13A3 kartiert werden (Fornoff F., unveröffentlichte Resultate). Exklusiv in Hodengewebe konnte ein 5´ verlängerter RACE-Klon erhalten werden, dessen 5´Ende in Position 30.280 im BAC RP13-13A3 kartiert wurde, was ein erster Hinweis für eine potentielle Gewebespezifität 5´ verlängerter *DDX3X* Transkripte war.

Bei Analyse der *DDX3X* TSS-Positionen (Tab. 4-4) aus der DBTSS-Sammlung wird festgestellt, daß von 29 ermittelten TSS-Positionen lediglich zwei TSSs die optimale Konsensussequenz für Initiatormotive enthalten (grün unterlegt in Tab. 4-4). Zehn weitere TSS-Positionen enthalten nur das Y-R-Dinukleotid und repräsentieren damit das minimale Inr-Motiv (grau unterlegt in Tab. 4-4). Für die restlichen 17 TSS-Positionen müssen daher Ungenauigkeiten bei der RLM-RACE Methode angenommen werden.

Für die weitere Kartierung der *DDX3X* TSS-Bereiche werden nur potentielle TSS mit Y-R-Dinukleotid oder Inr-Motiv aus Normalgeweben berücksichtigt. Es ist zu sehen, daß diese TSS-Positionen über einen Bereich von 815 bp, mit einer maximalen Distanz von 905 bp zum *DDX3X* ATG in Exon1, verteilt vorkommen. Dieser große Bereich weist auf die Anwesenheit mehrerer Core-Promoterregionen für *DDX3X* Transkripte hin, da eine einzelne Core-Promoter-region selten größer als 100 bp ist (Smale & Kadonaga, 2003; Sandelin et al., 2007; Juven-Gershon & Kadonaga, 2010). Die Forderung nach mehreren Core-Promoter Regionen für die *DDX3X* 5´ Region wird durch die Gewebeverteilung der TSSs in der DBTSS unterstützt (Tab. 4-4). Danach läßt sich die gesamte *DDX3X* TSS-Region in drei Subregionen mit unterschiedlicher Gewebeexpression unterteilen (I = ubiquitär; II = Hoden u. Hirn; III = nur Hoden).

Tab. 4-4: Position potentieller Initiator Core-Promotermotive für den *DDX3X* Transkriptstart.
Die drei TSS-Regionen sind durch rote Linien voneinander getrennt. Die Positionsangaben in Spalte 2 entsprechen der +1 Position (= TSS) des Transkriptes, die wiederum dem dritten Nukleotid in einem Inr-Motiv entspricht. Die Positionsangaben beziehen sich auf den Human X-BAC RP13-13A3. Mismatches der ermittelten Sequenzen zur Inr-Konsensussequenz sind in Spalte 3 rot markiert. Das hochkonservierte Y-R-Dinukleotid ist in Spalte 3 unterstrichen und in Spalte 4 durch Pluszeichen gekennzeichnet. Initiatormotive mit 100% Übereinstimmung zur Konsensus-sequenz sind im Fettdruck und grün hervorgehoben. Weitere potentielle TSSs mit assoziiertem Y-R-Dinukleotid sind hellgrau dargestellt. TSSs die nicht der Inr-Seq. entsprechen bzw. kein Y-R Dinukleotid enthalten, sind farblich nicht unterlegt. In Spalte 5 wird angegeben aus welchen Geweben/Zellen die Transkripte isoliert wurden. Die Zahlenangaben in Spalte 6 geben die Anzahl der Nennungen in der DBTSS an.

Position der 5´ Enden der DBTSS Klone			Sequenz des pot. Inr Elementes: YYANWYY	Y-R Regel für -1/+1 Position im Inr	Gewebe und Anzahl cDNA-Klone pro Gewebe	Gesamtzahl cDNA Klone in DBTSS
III		29.744	GCAACCG	+	Leukämie-Zellen	1
		29.755	GGAAGGT	-	Hoden	1
		29.824	TCCATTT	-	Hoden	2
		29.825	**CCATTTC**	+	Hoden	13
		29.827	ATTTCCC	-	Hoden	1
		29.828	TTTCCCC	-	Hoden	1
II	d	29.999	TCACAAA	+	Hirn	2
		30.077	CCACTAA	+	NT2	1
	c	30.191	AAAAAAA	-	Hirn	1
		30.204	AAACAAA	-	Hoden	1
		30.222	ACGAAAG	+	Hoden	1
	b	30.390	GAGAGCA	-	Uterus	1
		30.403	GTGTAGC	+	Trachea	1
		30.427	GCATTTC	+	Hirn	1
		30.428	CATTTCG	-	Hirn	1
	a	30.579	ACGTGCT	+	Testis	1
		30.608	GTATATT	+	Hirn (5), Testis (2), Darmkrebs	8
		30.609	TATATTA	-	Testis	1
		30.613	**TTAGATC**	+	Hirn	1
		30.633	TGCGCTC	-	Hirn	1
I		30.634	GCGCTCC	+	Trachea	1
		30.635	CGCTCCA	-	Neuroblastom	1
		30.637	CTCCAGA	-	Hirn, Thymus, Trachea, NT2	4
		30.638	TCCAGAG	-	Hirn (9), Trachea (4), Plazenta (2), Immunzellen, NT2, Zungentumor, Gliome	19
		30.639	CCAGAGC	+	Hirn (91), Trachea (33), Uterus (24), Testis (11), Thymus (10), Plazenta (8), Niere, Zunge (2), Darm, Milz, Arterien, HeLa, NT2 (16), Zungentumor (8), Lungentumor (7), Neuroblastom etc.	242
		30.640	CAGAGCC	-	Hirn (2), Niere, HeLa, Neuroblastom	6
		30.641	AGAGCCG	-	Nierentumor	2
		30.644	GCCGCAG	-	Hirn, Brusttumor	2
		30.677	GGAACAA	-	Hirn	1

Der erste, ATG-nahe TSS-Bereich (TSS-I), ist offensichtlich in allen Geweben aktiv. Die 5´ Grenze dieser TSS-Region wird in BAC-Position 30.634 gezogen, da alle weiteren TSS stromaufwärts auf Hoden- und Hirngewebe beschränkt sind (Bereich I, Tab. 4-4). Tabelle 4

zeigt deutlich, daß die ubiquitäre *DDX3X* Initiation in diesem Bereich sehr stark auf ein potentielles Initiator-Motiv „CCAGAGC“, in BAC-Position 30.639 (242 Einträge in DBTSS), fokussiert ist. Stromaufwärts ist ein potentielles TATA-Box-Motiv der Sequenz „TATATTAG“ in Position -32 relativ zur TSS lokalisiert. Unmittelbar stromabwärts (Abstand 0) der TATA-Box ist zudem ein potentielles BRE^{d} der Sequenz „ATCCGTG“ zu finden. Zu der Konsensussequenz „RTDKKKK“ sind zwei Abweichungen (rot markiert) zu verzeichnen. Stromabwärts des Initiator-Motives in BAC-Position 30.639 konnte kein DPE- oder MTE-Motiv in der geforderten Distanz identifiziert werden. Ein potentielles DPE-Motiv in der BAC-Position 30.647 mit der Sequenz „AGTTCT“, welches perfekt der DPE-Konsensussequenz entspricht, konnte keiner TSS aus der DBTSS-Datensammlung im geforderten Abstand von +28 zugeordnet werden. Weitere Core-Promoter-Motive konnten nicht im TSS-I Bereich von *DDX3X* kartiert werden. Somit besteht die *DDX3X* „Haupt-TSS-I“ vermutlich aus einem Core-Promotermodul mit drei Elementen: TATA-Box, BRE^{d}, Inr (Abb. 4-9).

Im Unterschied zur 73 bp langen ubiquitären *DDX3Y* TSS-I Region (BAC-Pos. 55.032 - 55.104), ist die *DDX3X* Initiationsregion somit wesentlich kürzer und auf 6 bp (BAC-Pos. 30.634 - 30.639) eingegrenzt. Die resultierenden *DDX3X* 5´UTRs der ubiquitären Transkripte haben damit Längen zwischen 91 und 96 bp.

Der mittlere TSS-Abschnitt (Bereich II, Tab. 4-4), wird auf Grund der Restriktion auf Hoden- und Hirngewebe, als **„T/B-TSS-II“** (T = Testis, B = Brain) von Position 29.999 bis 30.613 definiert. Nur acht der 14 TSS-Positionen, im 615 bp umfassenden Bereich II, entsprechen der minimalen Y-R-Regel. Anhand der „100 bp-Regel“ für eine Core-Promoterregion, ist es somit möglich den TSS-Bereich II weiter in vier Untergruppen, mit isolierten TSS-Positionen, zu unterteilen (Tab. 4-4). Die Subgruppe IIa umfaßt 35 bp, IIb 25 bp, IIc betrifft nur die BAC-Position 30.222 und IId umfaßt 79 bp. Eine klare Grenze mit unterschiedlicher Gewebespezifität der vier Untergruppen läßt sich nicht erkennen.
Die TSS in BAC-Position 30.613, Bereich IIa, liegt als einzige in einem perfekten Inr-Motiv (TTAGATC). An der nicht optimalen Position +29 relativ zu dieser TSS konnte hierzu ein potentielles DPE-Motiv mit einem Mismatch (AGCCGC) an Position 30.641 kartiert werden. Das Initiator-Motiv in BAC-Position 29.999 (TCACAAA; Bereich IId), welches das potentielle 5´Ende der TSS-II Region markiert, weist zwei Mismatches zur Konsensussequenz auf. Zusammen mit einer potentiellen TATA-Box „TATATTCA“ an der nicht optimalen Position -27 könnte ein Core-Promoter Modul gebildet werden. Für die weiteren sechs Y-R positiven TSS im *DDX3X* Bereich II konnten keine möglichen Assoziationen zu anderen allgemein bekannten Core-Promoter-Motiven festgestellt werden.

Für die TSS-Position 30.390 (Bereich IIb) könnte ein XCPE1-Motiv (X Core Promoter Element 1) ein alternatives Core-Promoterelement darstellen. Tokusumi et al. (2007) haben XCPE1 als Alternative zum Initiator-Motiv, in GC-reichen Core-Promoter Regionen beschrieben. Die Konsensussequenz lautet „DSGYGGRASM“, wobei ein Guanin oder Cytosin an Position neun (unterstrichen) als +1 Position, also als Transkriptionsstart, bezeichnet wird. Für die *DDX3X* TSS in BAC-Position 30.390 konnte die Sequenz „TTAGCGGAGA“ mit vier Abweichungen zur XCPE1-Konsensussequenz ermittelt werden. Die Autoren erwarten eine XCPE1-Häufigkeit von ~1% in Human Core-Promoterregionen proteinkodierender Gene. XCPE1 funktioniert nur in Kombination mit sequenzspezifischen Aktivator-TFs, wie z.B. NRF1, NF-1 und SP1, die an der Rekrutierung des PIC beteiligt sind. Alleine ist die durch XCPE1 gesteuerte Transkription auf dem Niveau der Background-Transkription (Tokusumi et al., 2007; Juven-Gershon et al., 2008). In Kapitel 4.6.2 wird gezeigt, daß die *DDX3X* 5´ Region (C,G)-reiche Sequenzabschnitte enthält, die ein solches XCPE1-Motiv funktionell begünstigen könnten.

Distal der Regionen I und II ist eine fokussierte *DDX3X* TSS in BAC-Position 29.825, in der DBTSS-Sammlung vorhanden, die nur in Hodengewebe identifiziert wurde (Bereich III, Tab. 4-4). Die fünf weiteren aufgeführten TSSs in diesem Bereich III entsprechen nicht den minimalen Core-Promoterkriterien. Stromaufwärts zur *DDX3X* TSS in BAC-Position 29.825, mit einer perfekten Initiator Sequenz „CCATTTC“, konnte in der relativen Position -28 ein nicht-kanonisches TATA-Box Motiv der Sequenz „CACCTGCG“, mit einem Mismatch, lokalisiert werden. Wie das TATA-Box Motiv stromaufwärts der *DDX3X* T/B-TSS-II, entspricht auch dieses Motiv nicht der bekannten AT-reichen TATA-Sequenz, ist aber konform mit den TATA-Box Regeln nach Smale & Kadonaga (2003; siehe Einleitung). Stromabwärts dieser TSS konnte ein potentielles DPE-Motiv „CCATCC“ in der optimalen Position +28 kartiert werden. Die DPE-Sequenz weist nur eine Abweichung (rot markiert) zur Konsensussequenz auf.

Aus den Daten der DBTSS wird auch deutlich, daß nicht alle kartierten *DDX3X* TSSs in gleichem Maße genutzt werden. Für viele der ermittelten TSSs ist lediglich ein cDNA-Klon registriert. Der TSS-I Bereich scheint auf eine Position (BAC-Position 30.639) konzentriert und der TSS-III Bereich ist ebenfalls fokussiert, da nur eine Position häufiger genutzt wird (BAC-Position 29.825). Die TSS-II Region erscheint laut der Klonhäufigkeit in DBTSS generell seltener genutzt zu werden. Die Kartierung der DBTSS-Daten unterstützt somit die Annahme von drei unterschiedlichen *DDX3X* TSS-Regionen.

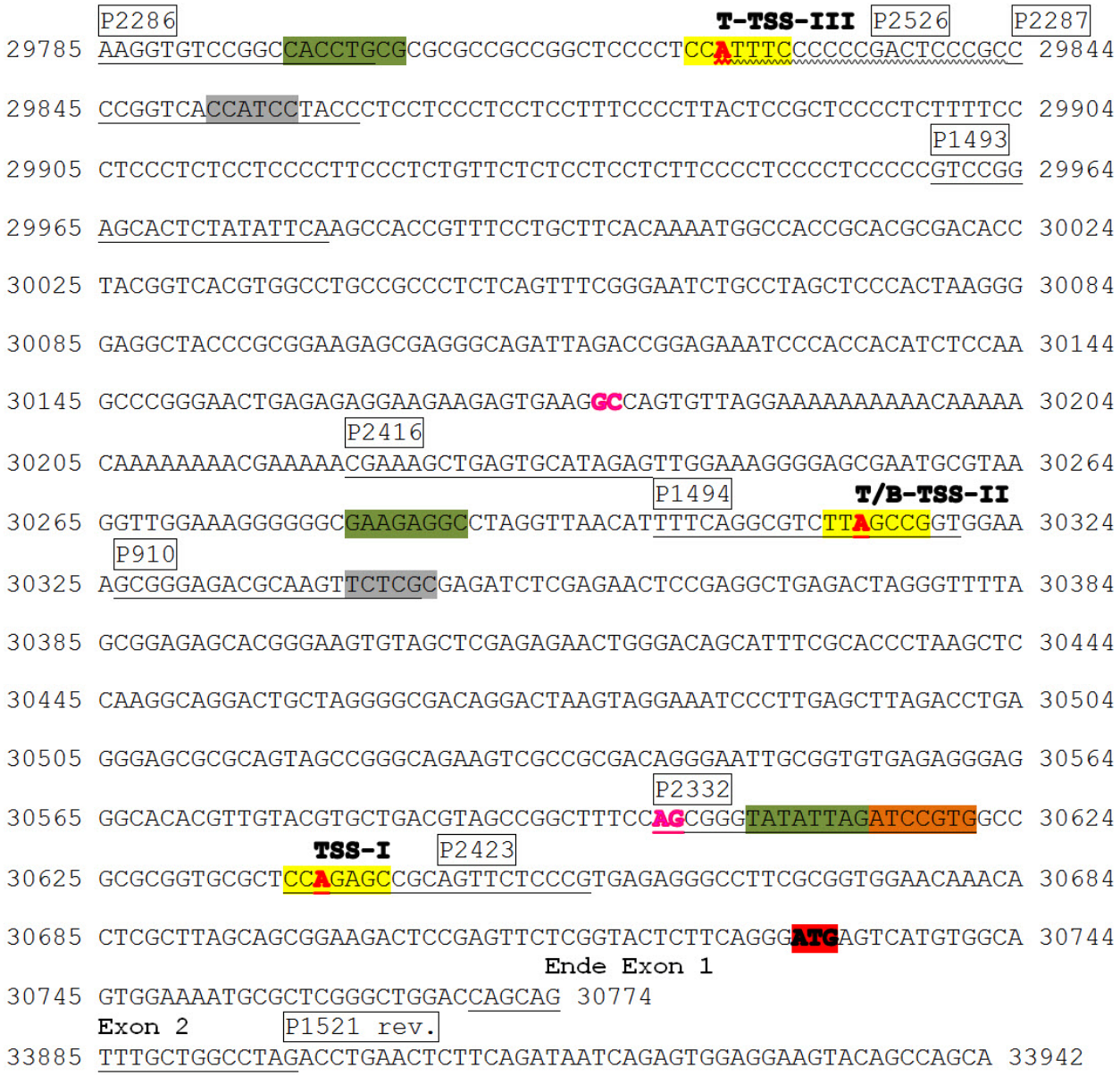


Abb. 4-9: Genomische Sequenz des *DDX3X* Genlocus von Position 29.785 bis 30.774 (Ende Exon1) im BAC RP13-13A3.

Die für die Analyse der Human *DDX3X* TSS-Regionen verwendeten Primer sind in der Sequenz unterstrichen. P2526 ist zur besseren Unterscheidung von P2287 gewellt unterstrichen. Der Reverseprimer P1521 ist überlappend in Exon1 und 2 eingezeichnet. Die identifizierten Core-Promoter Motive sind alle farblich hervorgehoben. Initiator-Motive sind gelb unterlegt und die TSSs innerhalb der Inr-Motive ist rot markiert. Potentielle TATA-Box-Motive sind grün, DPE-Motive sind hellgrau und das BRE^{d}-Motiv ist orange unterlegt. Das *DDX3X* Translationsstartcodon ATG ist rot hervorgehoben. Das Splicesignal GC-AG ist in Fettdruck und pink hervorgehoben. Die Primerbezeichnungen beziehen sich auf die interne Labor-Nomenklatur und sind mit einem Kasten dargestellt.

4.2.1 Kartierung der *DDX3X* Transkriptionsstartstelle I (TSS-I)

Zur experimentellen Überprüfung der DBTSS und Core-Promoter Analyse wurden gezielte RT-PCR-Analysen mit cDNAs aus Hodengewebe und männlichen und weiblichen Leukozyten durchgeführt. Die Leukozyten werden stellvertretend für die somatischen Gewebe genutzt, da bisher, außer in Hoden- und Hirngewebe, keine alternativen *DDX3X* 5´UTR-Varianten bekannt sind. Als Reverseprimer für alle *DDX3X* RT-PCR Experimente dient Primer P1521, der überlappend in Exon1 und Exon2 lokalisiert liegt.
Die von Fornoff (2003) erhaltenen 5´RACE-Klone aus Leukozyten und Hodengewebe mit Start in den BAC-Positionen 30.637 bzw. 30.642 sind, nach den DBTSS-Daten, dem ubiquitären *DDX3X* TSS-Bereich I zuzuordnen.

Mit Primer P2423 (BAC-Position 30.637), der 93 Nukleotide stromaufwärts des *DDX3X* ATGs lokalisiert ist (Primerlokalisation siehe Abb. 4-9), sind sowohl in Hodengewebe, als auch in männlichen und weiblichen Leukozyten RT-PCR Amplifikate erhältlich (Abb. 4-10). Primer P2332 ist 36 Nukleotide weiter stromaufwärts in BAC-Position 30.601 lokalisiert. PCR-Amplifikate mit P2332 sind nur in cDNA Proben aus Hodengewebe eindeutig zu beobachten. Mit cDNA Proben aus männlichen und weiblichen Leukozyten, konnten hingegen nur bedingt eindeutige PCR-Produkte gezeigt werden.

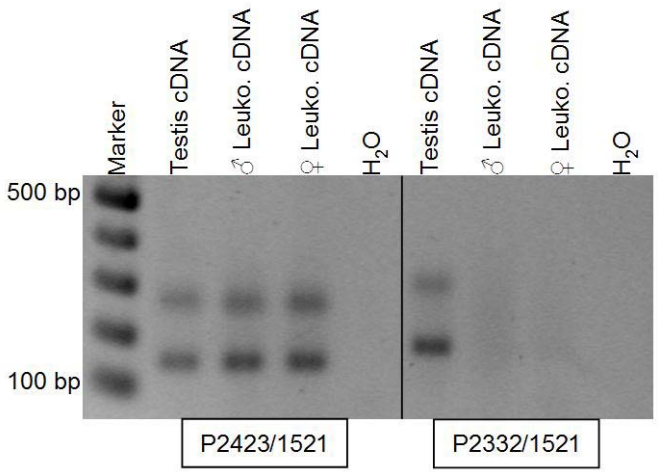


Abb. 4-10: RT-PCR Analyse der *DDX3X* TSS-I Region.
PCR-Produkt für P2423/1521 ist 151 bp, für P2332/1521 187 bp. Jeweils die untere PCR Bande ist das erwartete Amplifikat. Die obere Bande ist entweder ein Artefakt der PCR-Reaktion, oder verursacht durch Primerbindung an eines der zwei *DDX3* Pseudogene, oder an cDNAs anderer DEAD-Box Helikasen mit bisher unbekannter Sequenzhomologie. Das 4q24 *DDX3X* Pseudogen kann durch Primerpaar P2423/1521, mit einer Produktlänge von 147 bp amplifiziert werden. 35 PCR Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

In Abbildung 4-10 entspricht jeweils die untere PCR-Bande dem erwarteten Amplifikat. Die obere Bande könnten durch unspezifische Primerbindung an Transkripte von *DDX3* Pseudogenen, bzw. durch Artefaktbildung chimärer DNA-Moleküle nach Rekombination homologer Regionen in z.B. Pseudogentranskripten (Smyth et al., 2010; Schütze et al., 2011), verursacht werden. Es sind zwei *DDX3* Pseudogene in Xq13[21] und 4q24[22] lokalisiert (Kim et al.,

[21] *DDX3Y* Pseudogen in Xq13 befindet sich in BAC-Klon RP13-204A15, GenBank Acc.No. AL359740.24
[22] *DDX3X* Pseudogen in 4q24 befindet sich in BAC-Klon RP11-297P16, GenBank Acc.No. AC105460.4

2001; Chang & Liu, 2010). Eine Sequenzanalyse dieser unspezifischen PCR-Amplifikate wurde nicht durchgeführt, da der Fokus auf der Identifizierung der *DDX3X* TSS-Regionen liegt. Der ubiquitäre *DDX3X* TSS-I Bereich sollte somit zwischen den Positionen von Primer P2332 und P2423 zu suchen sein. Damit bestätigen die RT-PCR Ergebnisse die DBTSS und Core-Promoteranalyse und die ubiquitäre „***DDX3X* TSS-I**“ wird in BAC-Position 30.639 kartiert (Abb. 4-9).

4.2.2 Kartierung der selektiv genutzten *DDX3X* Initiationsstelle (T/B-TSS-II) im TSS-Bereich II

Durch die Analyse der *DDX3X* TSS in der DBTSS Datensammlung wurde gezeigt, daß ein zweiter TSS-Bereich von *DDX3X* existiert, der auf eine Expression in Hirn- und Hodengewebe beschränkt scheint (Tab. 4-4).

Zur Klärung der Gewebespezifität dieser 5´ verlängerten *DDX3X* Transkript-Variante wurde eine weitere Serie RT-PCRs mit den Primern P910 (BAC-Position 30.326), P1494 (BAC-Pos. 30.301) und P2416 (BAC-Pos. 30.221) durchgeführt, die distal von *DDX3X* TSS-I positioniert sind (Abb. 4-11).

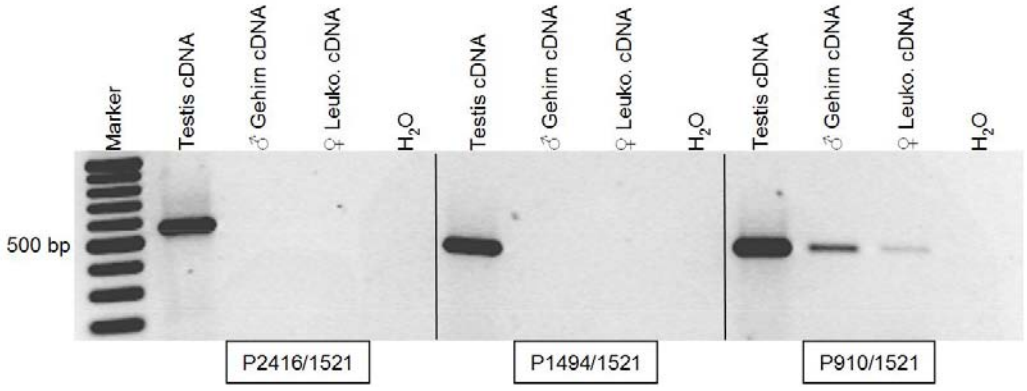


Abb. 4-11: RT-PCR Analyse zur Eingrenzung der selektiv genutzten *DDX3X* T/B-TSS-II.
PCR-Produktgrößen: P2416/1521 = 567 bp, P1494/1521 = 487 bp, P910/1521 = 462 bp. 35 PCR Zyklen. DNA Marker ist 100 bp DNA-Leiter. (Zimmer J., unveröffentlichte Resultate).

Die RT-PCR Ergebnissen mit Primerpaar P910/1521 zeigen deutlich PCR-Amplifikate mit cDNA-Proben aus männlichen Hoden- und Hirngeweben. Ebenfalls ist in Abb. 4-11 zu sehen, daß die *DDX3X* Transkription in Hodengewebe verstärkt, im Vergleich zu Hirngewebe, vorkommt. Mit cDNA-Proben aus weiblichen Leukozyten konnte ebenso eine minimale *DDX3X* Transkription festgestellt werden (Abb. 4-11). Dieses markiert einen Unterschied zu den vorangegangenen RT-PCR Experimenten mit cDNAs aus männlichen Leukozyten, in denen stromaufwärts von P2332 keine PCR-Amplifikate mehr detektiert werden konnten (Abb. 4-10).

PCR-Amplifikate mit den Primerpaaren P1494/1521 und P2416/1521 konnten jeweils nur noch in Hodengewebe festgestellt werden (Abb. 4-11). D.h. in männlichem Hirngewebe und in weiblichen Leukozyten findet offensichtlich keine *DDX3X* Transkription stromaufwärts der P910 Primerlokalisation (BAC-Position 30.326) statt. Diese Beobachtung konnte ebenso für weitere männliche und weibliche Gewebe gemacht werden. Eine Nutzung dieser Region als Start von *DDX3X* Transkripten ist, in männlichen Geweben, nur Hirn- und Hodengewebe vorbehalten (Zimmer J., unveröffentlichte Resultate).
Die mögliche Startstelle für den *DDX3X* T/B-TSS-II Bereich sollte also entsprechend der RT-PCR-Analyse zwischen den BAC-Positionen von P1494 (Pos. 30.301) und P910 (Pos. 30.326) zu finden sein.

Innerhalb dieser Sequenzregion wurde nur ein potentielles Initiator-Motiv, die Sequenz „TTAGCCG“ mit zwei Abweichungen zur Konsensussequenz, in BAC-Position 30.314, kartiert. Diese Position ist in der DBTSS Sammlung nicht vorhanden. Zu diesem Inr-Motiv konnte an Position -33 die Sequenz „GAAGAGGC“ lokalisiert werden, die ein potentielles TATA-Box Motiv darstellen könnte. Die gezeigte Sequenz entspricht allerdings nicht der klassischen AT-reichen TATA-Sequenz „TATAWAAR“, ist aber konform zu den TBP-Bindungsstudien von Smale & Kadonaga (2003; siehe Einleitung). In Kombination mit einem weiteren starken Core-Promoter Element, wie dem Initiator, ist es also durchaus möglich, daß TBP, durch die räumliche Struktur des TFIID-Komplexes, an diese Sequenz binden kann. Tokusumi et al. (2007) erwähnen, daß pro TATA-Motiv „TATAWAAR“ ein Mismatch toleriert werden könne. An Position 4 der Sequenz „GAAGAGGC“ ist ein Mismatch festzustellen. Die Autoren gehen allerdings nicht darauf ein, ob der Mismatch an jeder beliebigen Position auftreten darf, oder ob die sonst kritischen Positionen 2, 4 und 5 davon ausgenommen sind. Stromabwärts konnte ein potentielles DPE-Motiv „TCTCGC“ bestimmt werden, welches im geforderten Abstand von +28 lokalisiert ist. Die Sequenz entspricht der DPE-Konsensussequenz „DSWYVY“ nach Kutach & Kadonaga (2000).

Die durchgeführten RT-PCR Experimente im *DDX3X* T/B-TSS-II Bereich bestätigen somit nicht die Daten der DBTSS-Datenbank, die eine Hirn-/Hoden-TSS-Region bis zur Position 29.999 suggeriert haben. Auf Grund der RT-PCR Analyse, die keine PCR-Produkte mehr in Hirngewebe vor dem Primer P910 (BAC-Position 30.326) aufweist, wird die einzig vorhandene potentielle TSS in BAC-Position 30.314 als **„*DDX3X* T/B-TSS-II“** definiert. Als Core-Promoter-Modul für *DDX3X* T/B-TSS-II werden die potentiell funktionellen Motive TATA-Box, Initiator und DPE vorgeschlagen (Abb. 4-9). Weiter 5´ beginnende *DDX3X* Transkripte sind ausschließlich in Hodengewebe zu finden.

4.2.3 Kartierung der hodenspezifisch genutzten *DDX3X* Transkriptionsstartstelle III (T-TSS-III)

Aus der NCBI-Datenbank ist ein 5´ unvollständiges *DDX3X* Transkript (Acc.No. NM_001356.3) mit verlängerter 5´UTR bekannt, welches in BAC-Position 29.875 beginnt. Die zugehörige TSS könnte somit in BAC-Position 29.825 liegen (Tab. 4-4).

Um dies experimentell zu prüfen, wurde wiederum eine RT-PCR Analyse mit schrittweise 5´ verschobenen Primerpositionen (Abb. 4-9) durchgeführt.
Mit den stromaufwärts von P910 (BAC-Pos. 30.326) gelegenen Forwardprimern P1494 (BAC-Pos. 30.301), P2416 (BAC-Pos. 30.221), P1493 (BAC-Pos. 29.959), P2287 (BAC-Pos. 29.844) und P2526 (BAC-Pos. 29.825) konnten ausschließlich in Hodengewebe PCR-Amplifikate identifiziert werden (Abb. 4-11 & 4-12). Mit dem Forwardprimer P2286 (BAC-Pos. 29.785) konnte auch in Hodengewebe kein spezifisches *DDX3X* Produkt mehr gezeigt werden (Abb. 4-12). Die sichtbaren PCR-Amplifikate für P2286/1521 entsprechen weder in Größe noch in Sequenz den zu erwartenden *DDX3X* Transkripten.

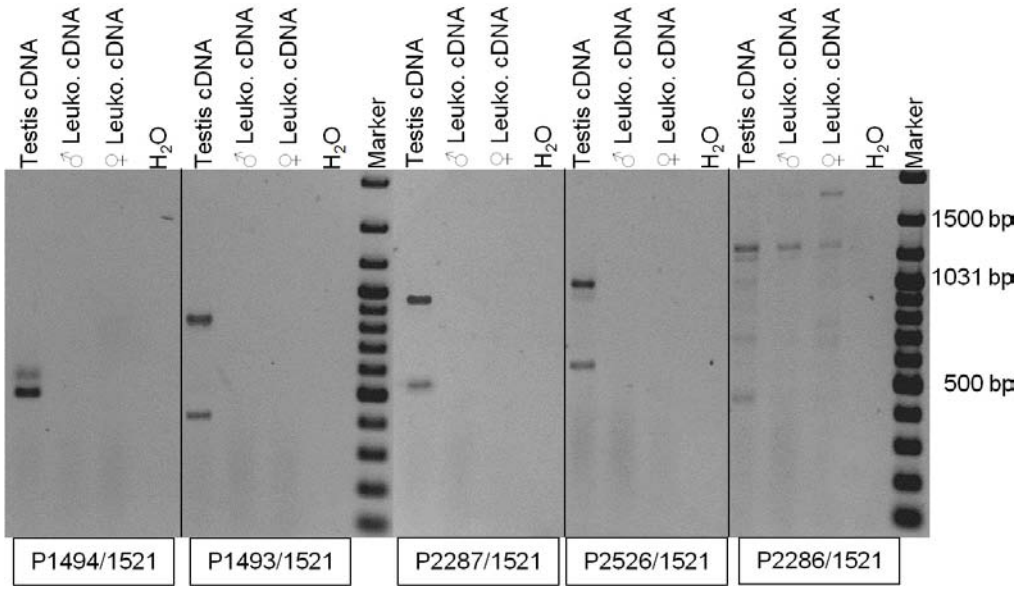


Abb. 4-12: RT-PCR Analyse *DDX3X* T-TSS-III Transkripte.
PCR-Produktgrößen für P1494/1521 = 487 bp, für P1493/1521 = 829 bp (403 bp), für P2287/1521 = 944 bp (518 bp), für P2526/1521 = 963 bp (537 bp) und für P2286/1521 = 1003 bp (577 bp). Werte in Klammern geben Größe des *DDX3X* Splicingproduktes an. Mit Primer P2286/1521 konnten keine spezifischen PCR Produkte erhalten werden, Primer bindet unspezifisch. Die sichtbaren Banden entsprechen nicht den errechneten Werten für *DDX3X* Transkripte. Das prozessierte *DDX3X* Pseudogen in 4q24 scheint ein Derivat des langen hodenspezifischen *DDX3X* Transkriptes (T-TSS-III) zu sein. Es ist bisher nicht bekannt, wo das 4q24 *DDX3X* Pseudogen initiiert. Im Alignment sind Homologien zur nativen 5´ *DDX3X* Sequenz erkennbar. Eine Primerbindung an ein mögliches Pseudogentranskript scheint möglich und würde für P2286/1521 ein PCR-Amplifikat von 1230 bp erzeugen. Dies könnte auch die Banden in somatischen Geweben erklären. Für P1494/1521 ist die obere, schwächere Bande ebenfalls durch ein mögliches *DDX3X* Pseudogentranskript mit 569 bp zu erklären. Wie auch in Abb. 4-10 sind auch chimäre RT-PCR-Produkte, durch Rekombination und Vervielfältigung homologer Sequenzen, möglich (Smyth et al., 2010; Schütze et al., 2011). 35 PCR Zyklen. DNA-Marker ist 100bp DNA-Leiter.

Der potentielle Transkriptstart dieser dritten *DDX3X* Transkriptklasse, mit hodenspezifischer Expression, sollte somit zwischen den Positionen 29.785 und 29.825 zu finden sein (Abb. 4-9). Die P2526 Primerposition entspricht der TSS-Position aus der DBTSS Analyse, so daß BAC-Position 29.825 als **„*DDX3X* T-TSS-III“** definiert werden kann. Transkripte mit Start in *DDX3X* T-TSS-III scheinen nicht nur hodenspezifisch zu sein, sondern auch spezifisch für das männliche Geschlecht. Spezifische RT-PCRs, mit Primern stromaufwärts zum Forwardprimer P910 (Abb. 4-9), mit weiblichen Humangeweben (Ovar, Niere, Leukozyten) ergaben keine positiven Resultate (Zimmer J., unveröffentlichte Resultate; nicht gezeigt).

Eine Besonderheit der hodenspezifischen *DDX3X* T-TSS-III Transkripte ist ein Splicing innerhalb der verlängerten *DDX3X* 5´UTR. Das Splicing wurde per Sequenzanalyse von Jakob (2005) zwischen den BAC-Positionen 30.177 und 30.602 (426 bp) bestimmt. Das *DDX3X* 5´UTR-Splicing ist nicht homolog zu den drei *DDX3Y* 5´UTR-Splicevarianten. Die gesplicte *DDX3X* 5´UTR hat ein Länge von 479 Nukleotiden. Das Donor-Splicesignal ist ein GC-Dinukleotid, welches zusammen mit dem AG-Akzeptorsplicesignal in nur 0,5 bis 1% aller Splicesignale vorkommt (Thanaraj & Clark, 2001; Churbanov et al., 2008). Das GC-AG Splicesignal, welches, wie auch das GT-AG Signal, vom U2-Spliceosom („major eukaryotic Spliceosome“) prozessiert wird, kommt vorwiegend in alternativen Exonen vor. Die Akkumulation von alternativen GC-Signalen in der Säugetier- und v. a. in der Humanevolution wurde vermutlich durch eine funktionelle Selektion für alternative Splicing-Prozesse getrieben. In Abb. 4-12 ist zu sehen, daß das *DDX3X* 5´UTR-Splicing mit den Forwardprimern P1493, P2278 und P2526 darstellbar ist. P1494 ist stromabwärts des Donor-Splicesignals gelegen und erfasst somit keine gesplicten RT-PCR Produkte (Abb. 4-12).

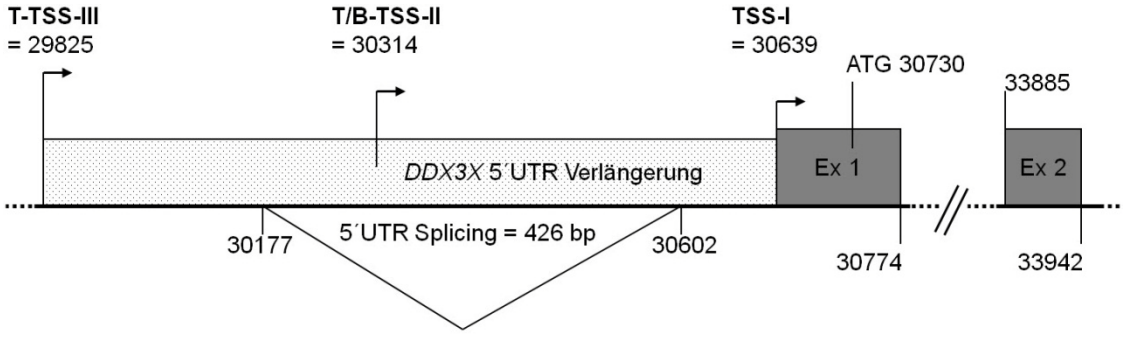


Abb. 4-13: Graphische Darstellung der drei *DDX3X* Transkriptionsinitiationsstellen.
Eingezeichnet sind die drei *DDX3X* TSSs mit ihrer Position im BAC RP13-13A3, sowie das 5´UTR Splicing der Transkripte mit Start in *DDX3X* T-TSS-III. Die 5´UTR Bereiche, die hodenspezifisch exprimiert werden, sind gepunktet eingezeichnet. Die ubiquitär exprimierten Exone1 und 2 sind ausgefüllt dargestellt.

Für die *DDX3X* Transkriptionskontrolle läßt sich anhand der RT-PCR Analyse und den Datenbankrecherchen also ein Model mit drei Core-Promoterregionen postulieren (Abb. 4-9 & 4-13). Insgesamt konnte eine gute Übereinstimmung der *in-silico* Daten mit den RT-PCR und

5´RACE Experimenten festgestellt werden. Der proximale Core-Promoter ist in seiner Expression nicht gewebespezifisch und ermöglicht eine ubiquitäre *DDX3X* Transkription. Als zweite *DDX3X* TSS-Region konnte T/B-TSS-II mit einer Restriktion auf Hoden- und Hirngewebe gezeigt werden. Für *DDX3X* T-TSS-III konnte ein für Hodengewebe spezifischer Core-Promoterbereich kartiert werden. Für alle drei *DDX3X* Initiationsstellen konnten Core-Promotermodule aus drei Motiven vorgeschlagen werden. In allen drei Fällen handelt es sich um eine Kombination aus TATA- und Initiator-Core-Promotermotiven.

4.3. Vergleich der kartierten *DDX3Y* und *DDX3X* TSS-Regionen

In diesem Abschnitt der *DDX3Y* und *DDX3X* Transkriptionsstartanalyse werden die zuvor beschriebenen TSS-Regionen beider Gene auf Gemeinsamkeiten und Unterschiede miteinander verglichen. Für diesen Sequenzvergleich wurde das Sequenzalignment-Programm ClustalW eingesetzt (Larkin et al., 2007). Um ein möglichst optimales Sequenzalignment zu garantieren, wurden verschiedene Parametereinstellungen getestet. Die daraus als optimal definierten Parameter (Gap-Open = 10, Gap-Extension = 1,0) ermöglichten eine optimale Alignmentqualität für den Erhalt zusammenhängender Sequenzblöcke, die in Kapitel 4.6.3.1 als Regionen mit konservierten TFBSs erkannt werden. Als Anker für das Sequenzalignment diente der hochkonservierte, kodierende Bereich des *DDX3Y*/*DDX3X* Exon1 (Abb. 4-14).

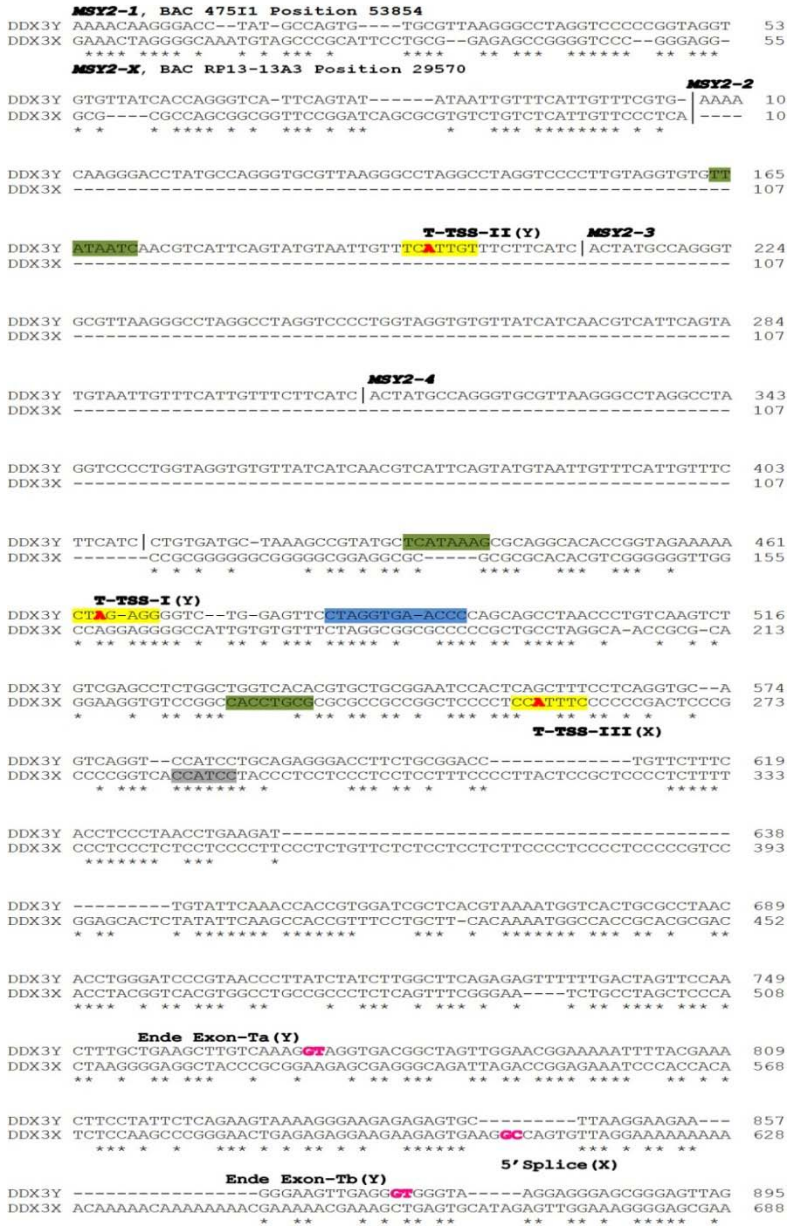
Durch das Sequenzalignment konnte stromaufwärts der *DDX3X* T-TSS-III Position eine 107 bp lange Sequenz (BAC RP13-13A3, Pos. 29.570 - 29.676) identifiziert werden, die eine deutliche Homologie zum *MSY2*-Repeat vor *DDX3Y* aufweist (Abb. 4-14). Diese X-chromosomale Sequenz mit einer Sequenzidentität von 63,4% zu *DDX3Y MSY2-1* wird als **„*MSY2-X*“** definiert. Eine detaillierte Funktionsanalyse der *MSY2* und *MSY2-X* Sequenzen erfolgt in Kapitel 4.10.
Der Vergleich der proximalen *DDX3Y* und *DDX3X* Promoterregionen (Abb. 4-14) zeigt deutlich, daß die ubiquitäre Genexpression in den TSS-I Core-Promoterregionen über unterschiedliche Initiationsmechanismen realisiert wird. Für *DDX3Y* wurde der TSS-I Bereich mit 73 Nukleotiden und mit vier co-dominanten Inr-Motiven als breiter Core-Promoter-Typus erkannt. Die Fokussierung auf eine dominante TSS-I Position in BAC-Position 30.639 für *DDX3X* beschreibt den gegenteiligen Core-Promoter-Typ. Juven-Gershon & Kadonaga (2010) bezeichnen die Unterschiede, im Transkriptionsmechanismus dieser zwei Core-Promotertypen, als fundamental.
Beide Sequenzen weisen dennoch einen hohen Grad an Sequenzhomologie im TSS-I Bereich (~90%) auf. Veränderungen der Sequenz sind vor allem durch den Neuerwerb und/oder Verlust von Core-Promotermotiven zu erklären. An homologer Position zum *DDX3X* TSS-I TATA-Box Motiv (Position 30.607) ist in der *DDX3Y* Sequenz das erste optimale TSS-Ia Inr-Motiv in BAC-

Position 55.030 zu finden. Das BRE^{d} Motiv in der *DDX3X* Sequenz ist nicht in der *DDX3Y* Sequenz erhalten. Positionshomolog zum dominanten *DDX3X* Inr-Motiv ist für *DDX3Y* ein TATA-Motiv zu sehen. Für die drei weiteren Inr-Motive und das DPE in der *DDX3Y* Sequenz ist auf *DDX3X* Seite keine Entsprechung zu finden.

Für den *DDX3Y* TSS-I-(T) Core-Promoter (BAC-Pos. 54.911) konnte experimentell als auch *in-silico*, bis auf ein Y-R Dinukleotid, kein vergleichbares *DDX3X* Core-Promotermodul festgestellt werden. Eine Kartierung von 5´Transkriptenden ist allerdings für diesen *DDX3X* Bereich per RT-PCR Analyse nicht zu erreichen, da auch stets längere *DDX3X* Transkripte, die von T/B-TSS-II oder T-TSS-III starten, amplifiziert werden.

Ein Vergleich der hodenspezifisch exprimierten *DDX3Y* und *DDX3X* Transkripte zeigt, daß für beide Gene gewebespezifische TSS vorhanden sind, die sich wohl seit der eingeschränkten Paarung der Gonosomen, vor ~ 165 - 310 Millionen Jahren (Lahn & Page, 1999; Veyrunes et al., 2008), jeweils Chromosomen-spezifisch weiterentwickelt haben.
So ist für die *DDX3X* T/B-TSS-II keine Entsprechung in der genomischen *DDX3Y* 5´ Sequenz zu finden ist (Abb. 4-14). Übertragen auf *DDX3Y* befindet sich die *DDX3X* T/B-TSS-II im Bereich von *DDX3Y* Intron-T. Mit der RT-PCR zu *DDX3Y* TSS-I-(T) konnte bereits gezeigt werden, daß im *DDX3Y* Intron-T keine Initiation stromaufwärts der TSS-I-(T) stattfindet.

Die Position der *DDX3X* T-TSS-III ist im Vergleich zur *DDX3Y* Sequenz im *DDX3Y* Exon-T lokalisiert, 96 Nukleotide stromabwärts zur kartierten *DDX3Y* T-TSS-I Position. Somit liegt die *DDX3Y* T-TSS-I Position bereits außerhalb der längsten *DDX3X* 5´UTR-Sequenz.
Für *DDX3Y* T-TSS-II mit Start im *MSY2* Minisatelliten-Repeat gibt es keine Entsprechung auf *DDX3X* Seite. Die hodenspezifische Initiationsstelle *DDX3Y* T-TSS-II befindet sich 362 Nukleotide weiter 5´ als die *DDX3X* T-TSS-III Position, welche die längsten hodenspezifischen *DDX3X* 5´UTR Transkripte darstellt.



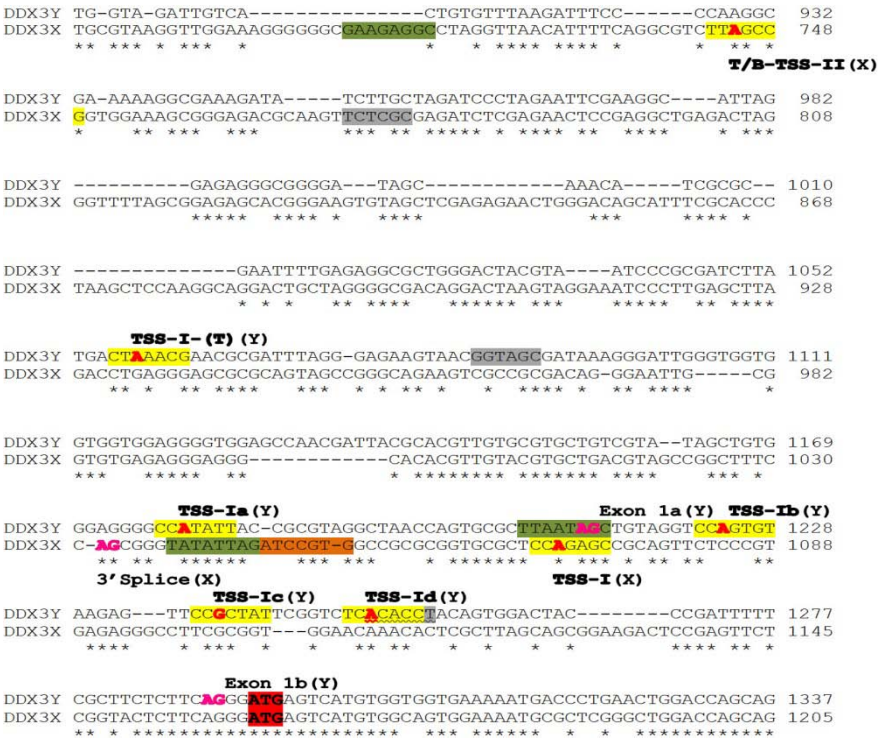


Abb. 4-14: Vergleich der genomischen *DDX3Y* und *DDX3X* 5´ Sequenzen.

Die *DDX3Y* Sequenz reicht von Anfang *MSY2* bis Ende Exon1 (BAC 475I1 Pos. 53.854 bis 55.190 = 1.337 bp), die *DDX3X* Sequenz beginnt in homologer Region zu *DDX3Y MSY2* mit *MSY2-X* und endet ebenfalls mit Exon1 (BAC RP13-13A3 Pos. 29.570 bis 30.774 = 1.205 bp). Die kartierten *DDX3Y* und *DDX3X* Inr-Motive sind gelb markiert, die TSSs sind mit roten Buchstaben A oder G hervorgehoben. Potentielle TATA-Box Motive sind grün, MTE-blau, BRE^{d}- orangefarben und DPE-Motive sind grau dargestellt. Das *DDX3Y* TSS-Id Inr-Motiv fällt mit einem DPE zusammen, welches durch eine gewellte Unterstreichung hervorgehoben wird. Das ATG Startcodon ist rot hervorgehoben, die 5´ und 3´ Splicesignale sind in Fettschrift und pink angegeben. Die vier Y-spezifischen *MSY2* Repeats sind durch vertikale Striche voneinander getrennt. Von der *MSY2-X* Sequenz ist keine Vervielfachung bekannt. Dieser Bereich ist im Alignment als Lücke angegeben. Sterne unterhalb des Alignments geben homologe Nukleotide an.

4.4. Evolution der *DDX3Y* Keimbahn Core-Promoterstruktur in verschiedenen Säugerspezies

In Kapitel 4.1 wurde eine deutliche Komplexität der *DDX3Y* Transkriptionsstartstellen in Humangewebe identifiziert. Die Fragestellung für diesen Abschnitt ist, ob die dargestellten hodenspezifischen Transkriptvarianten humanspezifisch exprimiert werden, oder ob sie bereits in anderen Primaten- und Säugerspezies ebenfalls per RT-PCR zu detektieren sind.

Dazu wurde das *DDX3Y* Expressionsmuster vergleichend in verschiedenen Geweben, sowie Leukozyten, von Schimpanse (*Pan troglodytes* = Ptro), Rhesusaffe (*Macaca mulatta* = Mmul) und Weißbüschelaffe (*Callithrix jacchus* = Cjac) studiert. Zur *in-silico* Analyse der Core-Promoterregionen konnten für diese drei Primaten die genomischen *DDX3Y* Sequenzen in den Datenbanken identifiziert werden. In Ermangelung von DBTSS-Daten für *DDX3X* und *DDX3Y* in diesen Primaten orientiert sich die experimentelle TSS-Analyse an den Ergebnissen, die für das Human *DDX3Y* Gen ermittelt werden konnten. Auf Grund der geringen Verfügbarkeit von Primatengewebe mußte allerdings auf Methoden mit hohem Materialaufwand, wie 5´- und 3´-RACE-Analysen, zur experimentellen Kartierung der TSS verzichtet werden.
Zu den bereits erwähnten Primatenspezies wurden als Außengruppen die zwei Säugetiere Maus (*Mus musculus* = Mmus) und Rind (*Bos taurus* = Btau) in diese vergleichende Studie integriert. Von Rind und Maus konnten genomische Sequenzen beider Gene in den NCBI- und Ensembl-Datenbanken identifiziert werden. Für Maus *Ddx3y* waren zudem TSS-Daten in der DBTSS verfügbar.

Die Primatengewebe wurden vom „Biomedical Primate Research Center“ BPRC (Rijswijk, Niederlande) und vom „Deutschem Primatenzentrum“ DPZ (Göttingen) zur Verfügung gestellt. Die Analyse von Leukozyten war bei *Pan* und *Callithrix* durch die Untersuchung an lymphoblastoiden Zellinien, also transformierten B-Zellen, möglich.

4.4.1 Evolution der *DDX3Y* Keimbahn Core-Promoterstruktur in Primaten

Die homologen *DDX3Y* Sequenzen konnten per BLAST-Analyse und ClustalW-Sequenzalignment für *Pan troglodytes* im BAC-Klon „CH251-128L22“[23], für *Macaca mulatta* im BAC-Klon „CH250-541A11“[24] und für *Callithrix jacchus* im BAC-Klon „CH259-161K20“[25] lokalisiert werden (Abb. 4-15). Bei diesen *DDX3Y* Primatengenen handelt es sich um

[23] Ptro: GenBank Acc.No. AC146254.2; revers-komplementär.
[24] Mmul: GenBank Acc.No. AC213321.3; revers-komplementär.
[25] Cjac: GenBank Acc.No. AC225609.4; revers-komplementär.

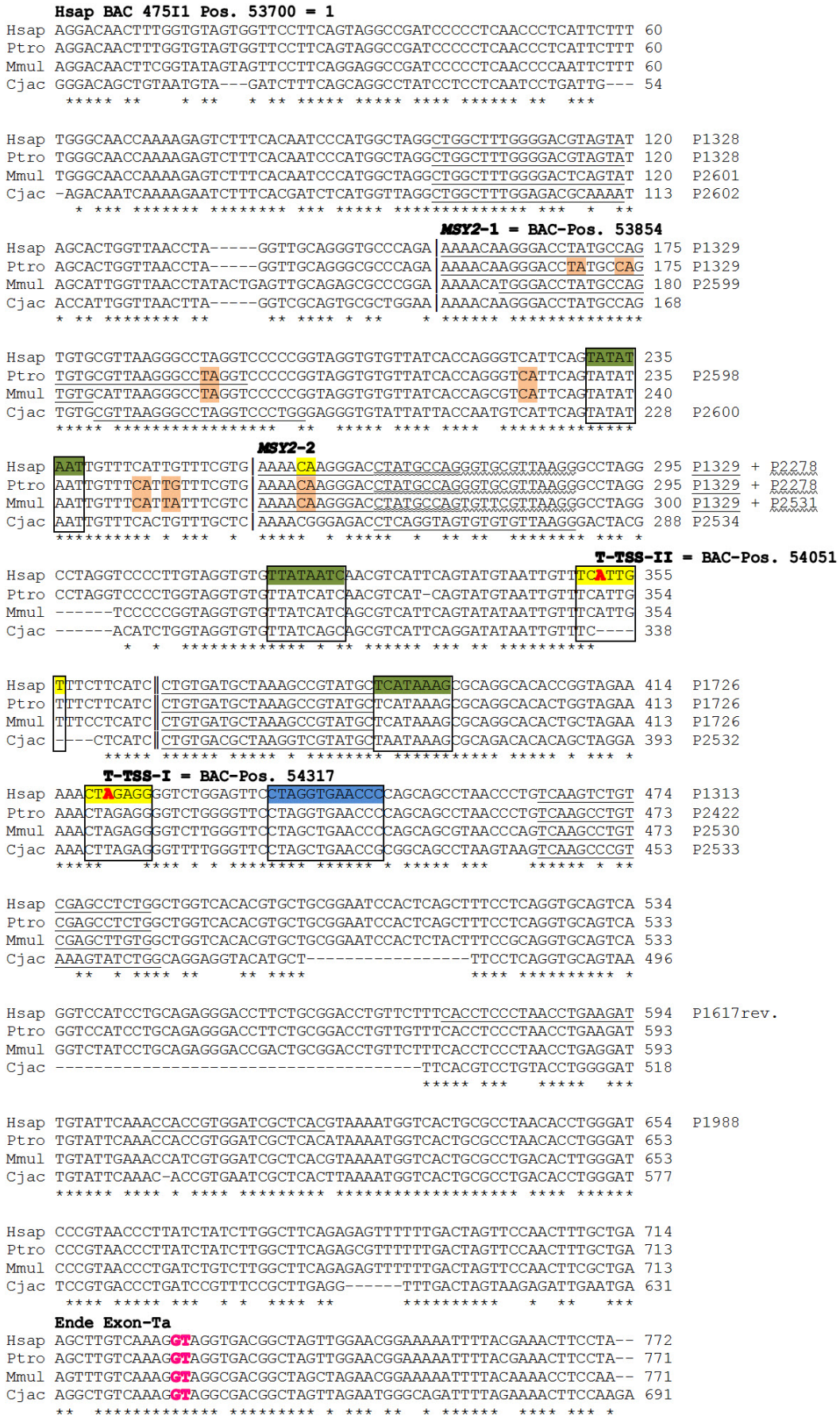
vollständige Gene mit identischer Anzahl von 17 Exonen, Translationsstart- und Stopcodon und vorhandenen Polyadenylierungsstellen. Die Homologien in den kodierenden Primaten *DDX3Y* Sequenzen betragen zwischen 94 und 99%. Die Tabellen mit den Primaten *DDX3Y* Exon-Intronstrukturen befinden sich im Anhang (Tab. A-4 bis A-7).

Im *DDX3Y* Sequenzalignment ist zu sehen, dass bei allen untersuchten Primaten zwei *MSY2* Repeats (*MSY2-1* und *-2*) in den potentiellen 5´UTR-Sequenzen vorhanden sind (Abb. 4-15). Bao et al. (2000) hatten dies bereits zuvor, über genomische PCR-Analysen, für Schimpanse (GenBank Acc.No. AF207845.1) und Orang-Utan (GenBank Acc.No. AF207846.1) gezeigt.

In der vorliegenden Analyse konnte somit zusätzlich gezeigt werden, daß stromaufwärts des *DDX3Y* Gens bei *M. mulatta* und beim Neuweltaffen *C. jacchus* zwei *MSY2* Repeateinheiten lokalisiert sind (Rauschendorf et al., 2011). Ein detaillierter *MSY2*-Sequenzvergleich zwischen Mensch und den drei untersuchten Primatenarten ist im Kapitel 4.10 zu finden.

Der Vergleich der *H. sapiens DDX3Y* TSS-I Region, mit den definierten Grenzen TSS-Ia und TSS-Id, zeigt, daß in den homologen Primatenregionen genomisch ebenfalls mehrere potentielle Initiationsstellen mit Inr-Motiven vorhanden sind (mit Kästen in Abb. 4-15 markiert). Y-R-Dinukleotide in den TSS-Ia und -Id Inr-Motiven sind in allen vier Spezies vorhanden. Einzig bei *M. mulatta* sind Y-R Punktmutationen, in den zwei mittleren Inr-Motiven der TSS-I Region, festzustellen. Da für *H. sapiens* die ganze Region bereits experimentell als breiter Core-Promoter identifiziert wurde, ist somit wahrscheinlich, daß auch bei den anderen Primaten eine breite Core-Promoterregion, mit mehreren Initiationsstellen, für *DDX3Y* TSS-I vorliegt.
Das potentielle TATA-Motiv der Human *DDX3Y* TSS-I Core-Promoter Region (BAC-Pos. 55060; Kombination mit Inr-Motiv TSS-Ic) ist bei *C. jacchus* deletiert und nicht mehr funktionell (Abb. 4-15). Das potentielle DPE-Motiv (BAC-Position 55.104; Kombination mit Inr-Motiv TSS-Ib), welches mit dem Human TSS-Id Inr-Motiv überlappt, ist bei *Callithrix* durch „indel“-Ereignisse (Insertion & Deletion) nicht mehr erkennbar.

Für die experimentell evaluierte Human *DDX3Y* TSS-I-(T) in BAC-Position 54.911 konnte in dem Primaten-Alignment ein Inr-Motiv an homologer Position festgestellt werden (Abb. 4-15). In *Macaca* entspricht das Inr-Motiv „CTAAACC“ einer perfekten Inr-Konsensussequenz. In den drei anderen Spezies ist jeweils ein Mismatch in der letzten Position des Inr-Motives vorhanden. Das potentiell assoziierte DPE-Motiv „GGTAGC“ (BAC-Position 54.939) ist ebenso in den genomischen Sequenzen von Ptro, Mmul und Cjac wiederzufinden. In *Macaca* und *Callithrix* weist das DPE-Motiv jeweils zwei Abweichungen zur DPE-Konsensussequenz auf („GATAGC“ bzw. „CGTACC“). Dies ist jeweils ein Mismatch mehr als in *Homo* und *Pan* („GGTAGC“).


Hsap BAC 475I1 Pos. 53700 = 1
MSY2-1 = BAC-Pos. 53854
MSY2-2
T-TSS-II = BAC-Pos. 54051
T-TSS-I = BAC-Pos. 54317
Ende Exon-Ta

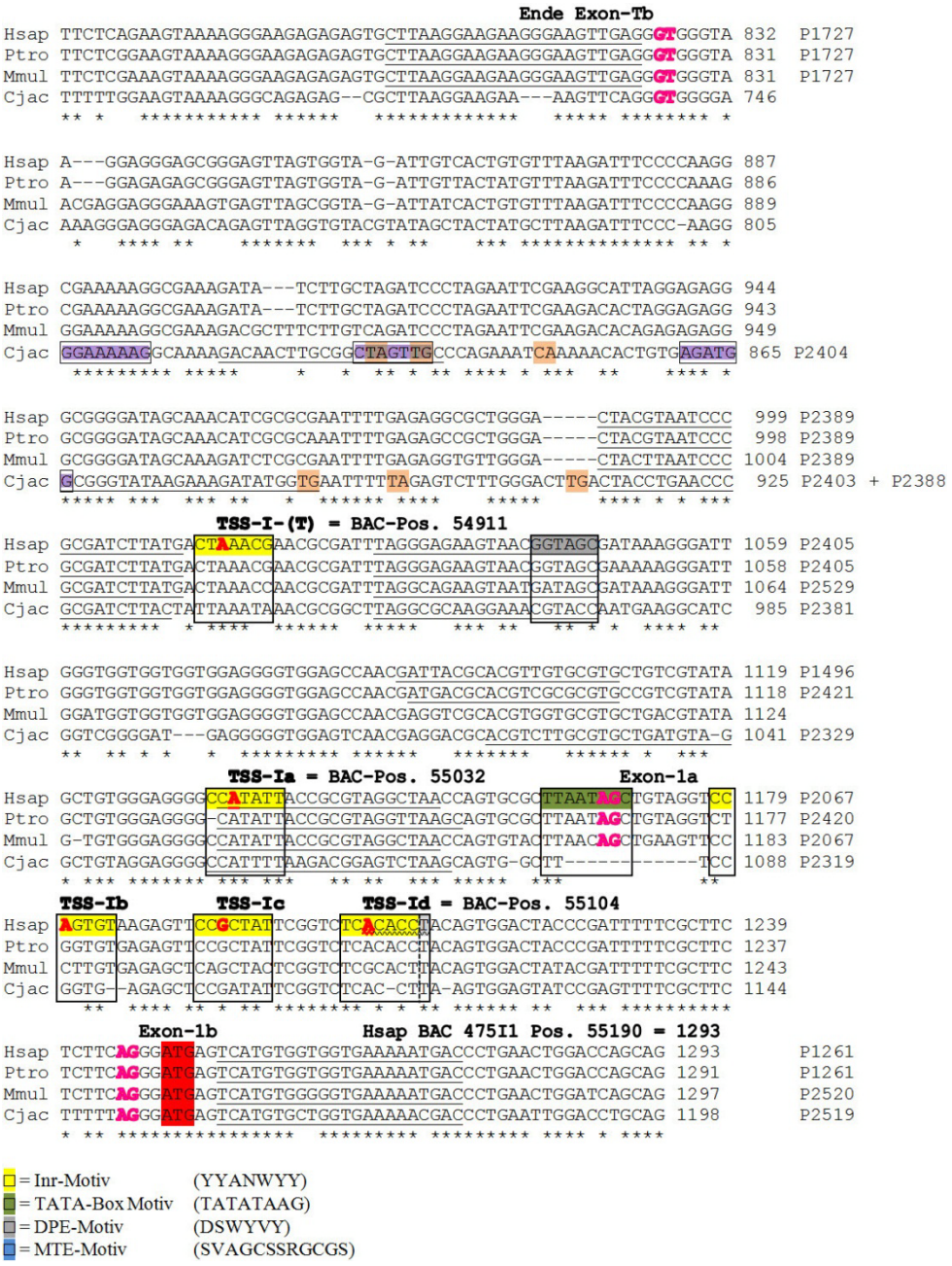


Abb. 4-15: Genomisches *DDX3Y* Sequenzalignment zwischen Mensch (Hsap), Schimpanse (Ptro), Rhesusaffe (Mmul) und Weißbüschelaffe (Cjac).

Humansequenz vor *MSY2* bis Ende Exon 1 (Pos. 53.700 bis 55.190 in BAC 475I1 = Pos. 1 - 1293 in Alignment). Ptro BAC-Pos. 60.270 – 58.980 (CH251-128L22), Mmul BAC-Pos. 84.244 – 82.948 (CH250-541A11) und Cjac BAC-Pos. 52.803 – 51.606 (CH259-161K20). Kartierte Core-Promotermotive in der Hsap *DDX3Y* Sequenz wie in Abb. 4-2. Abweichende Inr-Motive bzw. Y-R-Dinukleotide in Non-Human Primaten sind orange dargestellt. Ein pot. Core-Promotermodul für Cjac TSS-I-(T) ist gerahmt und lila dargestellt. Erläuterung siehe Text. Inr-Motive gelb markiert, TSSs mit roten Buchstaben A und G hervorgehoben. TATA-Box Motive sind grün, MTE- hellblau und DPE-Motive sind hellgrau dargestellt. Die potentiell positionshomologen Core-Promotermotive in den vier Sequenzen sind gerahmt dargestellt. Das *DDX3Y* TSS-Id Inr-Motiv fällt mit einem DPE zusammen, welches gewellt unterstrichen dargestellt ist. Ende des TSS-Id Inr-Motives durch gestrichelte Linie angezeigt. Das ATG Startcodon ist rot hervorgehoben, die 5´ und 3´Splice-Signale sind in Fettschrift und pink hervorgehoben. Die Y-spezifischen *MSY2*-Repeats sind durch vertikale Striche voneinander getrennt. Die Human *MSY2*-Repeateinheiten 3 und 4 sind entfernt und durch vertikalen Doppelstrich markiert. P2278 bzw. P2531 in *MSY2-2* ist gewellt unterstrichen. Im Überlappungsbereich mit P1329 doppelt gewellt unterstrichen. Exone 2 und 3 sind nicht gezeigt. Reverseprimer für Ptro (P2419), Mmul (P1286) und Cjac (P2513) sind positionshomolog bzw. identisch zu Hsap Reverseprimer P1286. Sterne unterhalb des Alignments geben homologe Nukleotide an.

DDX3Y TSS-I RT-PCR Analyse:

Pan troglodytes:

Die RT-PCR Analyse für *P. troglodytes* ergab mit dem Primerpaar P2420/2419 in den überprüften Geweben eine ubiquitäre *DDX3Y* Expression (Abb. 4-16). Eine Schwächung der *DDX3Y* Expression ist ab Primer P2421/2419, besonders mit dem Primerpaar 2405/2419, in somatischen Geweben festzustellen. Eine deutliche Expression ist nur noch in Hoden- und abgeschwächt in Nierengewebe festzustellen. In Leber- und Milzgewebe, sowie in B-Zellen des Schimpansen ist keine eindeutige Expression mehr erkennbar. Das negative Ergebnis mit P2389/2419 zeigt, daß, wie für auch den Menschen, beim Schimpansen eine *DDX3Y* Transkriptionsstartstelle zwischen den Primerpositionen P2405 und P2389 lokalisiert sein muß. D.h. die Expressionsmuster von Ptro *DDX3Y* sind im Bereich TSS-I ubiquitär und werden im Bereich von TSS-I-(T) nur noch stark in Hodengewebe exprimiert. Dieser Befund ist vergleichbar mit dem Human-Ergebnis, daß eine *DDX3Y* 5´ Verlängerung im TSS-I Bereich vorhanden ist und diese vornehmlich, aber nicht exklusiv, in Hodengewebe transkribiert wird.

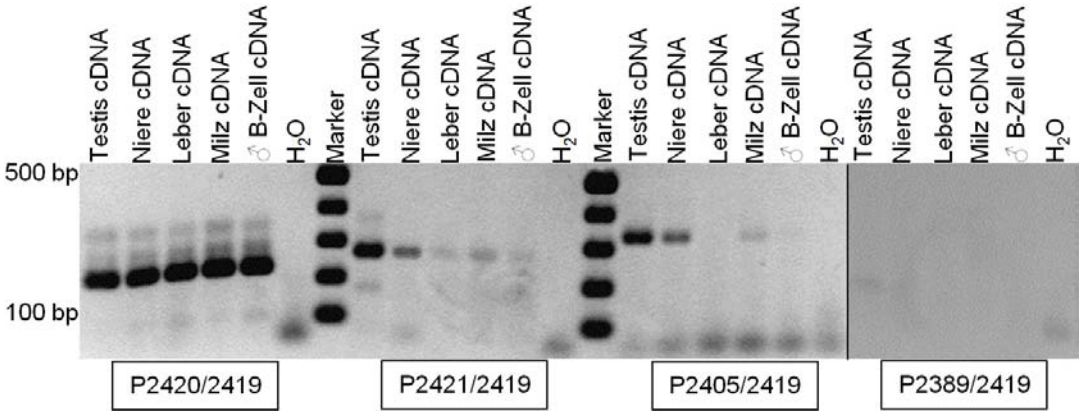


Abb. 4-16: RT-PCR Analyse von *Pan troglodytes DDX3Y* TSS-I Region.
P2420/2419 PCR-Amplifikat ist 223 bp groß, P2421/2419 ist 265 bp groß und P2405/2419 ist 328 bp groß, errechnete Produktgröße für P2389/2419 ist 368 bp. Agarosegel ist im Bereich von P2420/2419 schräg im elektr. Feld gelaufen. 35 PCR-Zyklen. DNA-Marker ist 100bp DNA-Leiter.

Macaca mulatta:

Bei der vergleichbaren RT-PCR Analyse mit *Macaca mulatta* Geweben konnten mit P2067/1286 in allen untersuchten Geweben, außer in männlichen Leukozyten, PCR-Amplifikate festgestellt werden (Abb. 4-17). RT-PCRs mit der identischen Leukozyten cDNA ergaben positive Resultate für Mmul *DDX3X* und die Mmul β-Aktin Kontrolle (Abb. 4-34). Eine qualitativ schlechte cDNA kann somit ausgeschlossen werden. Eine zusätzliche RT-PCR Überprüfung mit P2520, hinter ATG in Mmul Exon1 lokalisiert, ergab ebenfalls keine Expression in Leukozyten (nicht gezeigt). Die 5´ Verlängerung bis zur *DDX3Y* TSS-I-(T) (mit P2529/1286 analysiert) scheint bei Mmul wesentlich stringenter auf Hodengewebe beschränkt zu sein, als

dies bei den Homininae Mensch und Schimpanse der Fall ist (Abb. 4-17). Mit P2389/1286 konnte in *Macaca* Hodengewebe noch ein PCR-Produkt dargestellt werden. Dieser Primer ist stromaufwärts der kartierten Human TSS-I-(T) lokalisiert. Es deutet also darauf hin, daß in *Macaca* eine weitere TSS existiert, die hodenspezifisch aktivierbar ist und stromaufwärts der vergleichbaren Human TSS-I-(T) Position lokalisiert ist (Abb. 4-17).

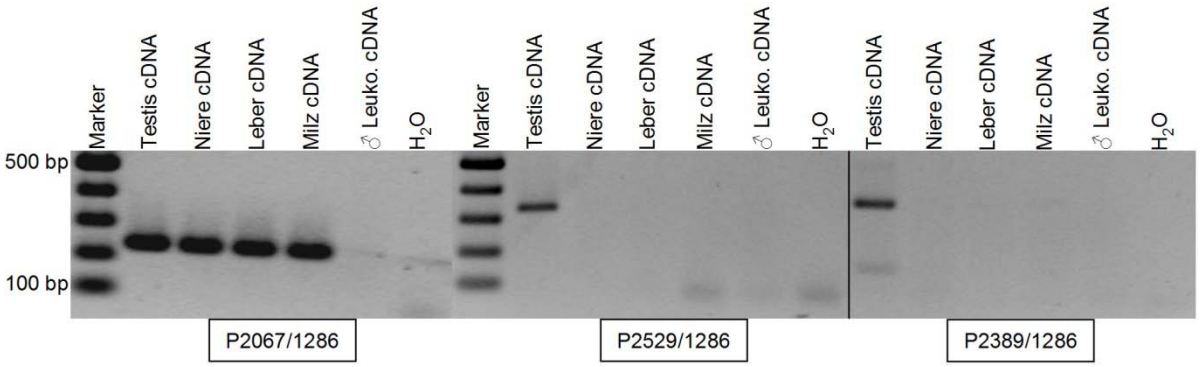


Abb. 4-17: RT-PCR Analyse *Macaca mulatta DDX3Y* TSS-I Region.
PCR-Amplifikatgröße beträgt für P2067/1286 225 bp, 330 bp für P2529/1286 und 370 bp für P2389/1286. 35 PCR-Zyklen. DNA-Marker ist 100 bp DNA-Leiter.

Callithrix jacchus:

Die RT-PCR Analyse in Abb. 4-18 zeigt, daß auch für *Callithrix jacchus* eine ubiquitäre *DDX3Y* Transkription im Bereich von TSS-I vorzufinden ist. Aber im Unterschied zum Menschen ist die ubiquitäre *DDX3Y* TSS-I Region bei *Callithrix* deutlich nach 5´ verschoben und endet nicht in der TSS-Ia Position (P2329 in Abb. 4-18).

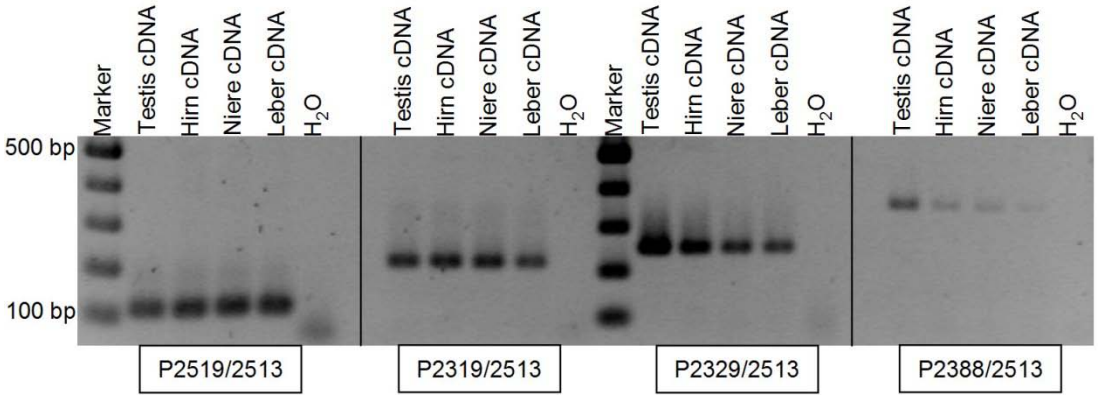


Abb. 4-18: RT-PCR Analyse *Callithrix jacchus DDX3Y* TSS-I Region.
PCR Amplifikatgrößen: P2519/2513 = 108 bp, P2319/2513 = 212 bp, P2329/2513 = 245 bp, P2388/2513 = 352 bp. 35 PCR-Zyklen. DNA-Marker ist 100 bp DNA-Leiter.

Ein zweiter deutlicher Unterschied bei *C. jacchus* ist, wie bei *Macaca*, eine weitere Verlängerung der *DDX3Y* Transkripte über die TSS-I-(T) Position hinaus (P2388 und P2403 in Abb. 4-18 & 4-19). Eine potentielle TSS ist zwischen den Primerpositionen P2403 und P2404 zu suchen, da mit P2404 keine PCR-Amplifikate mehr zu beobachten sind (Abb. 4-19). Im

Sequenzalignment (Abb. 4-15) sind zwischen diesen Primerpositionen mehrere potentielle Inr-Motive (CTAGTTG, TTGCCCA, TCAAAAA, GTGAATT, TTAGAGT, TTGACTA) bzw. Y-R-Nukleotide (in Abb. 15 markiert) feststellbar. Ein geeigneter Kandidat als Startposition dieses verlängerten Cjac *DDX3Y* Transkriptes wäre das erste Inr-Motiv „CTAGTTG“, da stromaufwärts in Position -28 eine pot. TATA-Box „GGAAAAAG“ und stromabwärts in -28 ein pot. DPE-Motiv „AGATGG“ markiert werden können (Abb. 4-15). Diese über TSS-I-(T) hinaus verlängerten *DDX3Y* Transkripte zeigen eine präferentielle Transkription in *Callithrix* Hodengewebe.

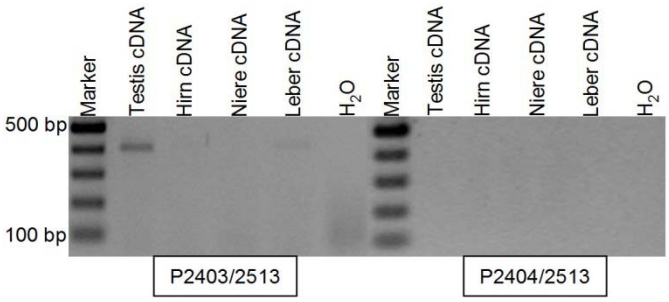


Abb. 4-19: *Callithrix jacchus DDX3Y* RT-PCR Analyse im Bereich von TSS-I-(T).
Berechnete RT-PCR-Amplifikatgrößen: P2403/2513 = 400 bp, P2404/2513 = 446 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100 bp DNA-Leiter.

Mit cDNA Proben aus einer männlichen, lymphoblastoiden *C. jacchus* Zellinie (hier nicht gezeigt) konnten mit allen Primern, außer P2403 und P2404, PCR-Amplifikate erhalten werden. Das Expressionsmuster ist vergleichbar zu *Callithrix* Hirn- und Nierengewebe.

Zusammenfassend kann also für den *DDX3Y* TSS-I Bereich gesagt werden, daß diese, in den vier untersuchten Primatenspezies die Initiationsregion für ubiquitäre *DDX3Y* Transkription repräsentiert. Anhand der Ergebnisse der RT-PCRs und den vergleichenden Core-Promoter Kartierungen kann für die 5´Grenze der ubiquitären Transkription innerhalb der Catarrhini (Hsap, Ptro, Mmul) die Position von *DDX3Y* TSS-Ia angenommen werden. Ausnahmen bilden die *Macaca* Leukozyten, in denen keine Expression festgestellt werden konnte, und der Neuweltaffe *Callithrix jacchus*, bei dem eine Verlagerung dieser Grenze über die Position von TSS-I-(T) hinaus zu beobachten ist. Für die Hominiden (Hsap, Ptro) konnte eine einheitliche 5´ Transkriptverlängerung mit präferentieller Expression in Hodengewebe bis zur TSS-I-(T) Position festgestellt werden. Für *Macaca* und *Callithrix* konnte eine nochmals längere 5´ Transkriptverlängerung von maximal 74 Nukleotiden (Position von Cjac P2403) festgestellt werden, die in *Macaca* exklusiv in Hodengewebe und in *Callithrix* präferentiell aber nicht exklusiv in Hodengewebe genutzt wird.
Diese Analyse verdeutlicht, daß sich, innerhalb der Hominiden *Homo* und *Pan*, das *DDX3Y* TSS-I und TSS-I-(T) Expressionsmuster einheitlich darstellt. Bei den beiden Primatenspezies *Macaca* und *Callithrix* wird eine deutliche Verlängerung der 5´UTR in *DDX3Y* Transkripten sichtbar, was für eine speziesspezifische und flexible Nutzung der vorhandenen Core-Promotermotive, im Kontext eines breiten Core-Promotertypus, spricht.

DDX3Y T-TSS-I & T-TSS-II RT-PCR Analyse:

Die Initiationsstelle T-TSS-I in *DDX3Y* Exon-T (Hsap BAC-Position 54.317) ist in den vier Primatenspezies genomisch an homologer Position zu finden (Abb. 4-15). Das Inr-Motiv „CTAGAGG“ ist bei den drei Catarrhini identisch konserviert. Für *Callithrix* ist eine Mutation im positionshomologen Y-R Dinukleotid von TA nach TT festzustellen, die somit das Inr-Motiv aufhebt. Alternativ kann bei *Callithrix* das Y-R Dinukleotid um eine Position verschoben werden, um wieder ein funktionelles TA-Dinukleotid zu erhalten. Das stromaufwärts gelegene TATA-Box Motiv „TCATAAAG“ ist wiederum in allen drei Catarrhini unverändert vorzufinden. Für *Callithrix* ist im TATA-Motiv „TAATAAAG“ eine positive Mutation in der kritischen Position 2 von C zu A zu finden. Es ist somit möglich, daß das stärkere TATA-Box Motiv bei *Callithrix* den vermeintlichen Verlust des Inr-Motives kompensieren kann und die Positionierung des PIC alleine übernimmt. Das stromabwärts gelegene MTE-Motiv „CTAGGTGAACCC“ ist in allen vier Primatenspezies positionshomolog kartierbar. Für *Macaca* („CTAGCTGAACCC“) und *Callithrix* („CTAGCTGAACCG“) konnte in Position 5 des MTEs, im Unterschied zu *Homo* und *Pan*, ein kanonisches Cytosin festgestellt werden. Anhand dieser *in-silico* Daten konnte also festgestellt werden, daß die Möglichkeit zur Transkriptionsinitiation im *DDX3Y* Exon-T in Position T-TSS-I auch in den anderen Primatenspezies existiert. Ob diese *DDX3Y* T-TSS-I Core-Promoter Motive in den anderen Primaten genutzt werden, muß die RT-PCR Analyse zeigen.

Die kartierte Human *DDX3Y* T-TSS-II Initiationsstelle (Hsap BAC-Position 54.051; *MSY2* Repeat 2; Inr-Block VI) ist in den genomischen Sequenzen von *Pan* und *Macaca* ebenfalls unverändert in homologer Position zu finden (Abb. 4-15 & 4-21). In der *DDX3Y* 5´ Sequenz des Neuwelt-affen *Callithrix* konnte das Inr-Motiv für T-TSS-II nicht identifiziert werden. In der *MSY2* Repeat-einheit 2 der *Callithrix jacchus* Sequenz fehlt, im Vergleich zu den drei Catarrhini, eine 8 bp lange Sequenz „ATTGTTTC“, die bei den Catarrhini verdoppelt vorliegt (Abb. 4-21). In dieser Sequenz befindet sich das *DDX3Y* T-TSS-II Inr-Motiv. Das stromaufwärts gelegene TATA-Motiv „TTATAATC“ (in Inr-Block IV) ist in allen untersuchten Primaten positionshomolog wiederzufinden. Allerdings ist nur in der Human Sequenz in der kritischen TATA-Box Position 5 ein kanonisches Adenin vorhanden (Abb. 4-15 & 4-21). Bei *Callithrix* ist zudem noch eine Veränderung in der TATA-Box Position 7 von T zu G festzustellen. Da in den nicht-human Primaten nur zwei *MSY2* Repeateinheiten vorliegen, könnte auch das kanonische TATA-Motiv „TATATAAT“, am 3´ Ende der Repeateinheit 1 (zwischen Inr-Block V und VI), in den anderen Primaten genutzt werden (Abb. 4-20). Das TATA-Motiv ist in allen vier Spezies zu 100% konserviert. Das stromabwärts gelegene Y-R Dinukleotid (Hsap BAC-Position 53.960), am 5´ Ende von Repeateinheit 2, ist in den vier Spezies vorhanden. Für die Catarrhini konnte ein CA-Dinukleotid und für *Callithrix* ein CG-Dinukleotid kartiert werden (Abb. 4-21). Stromabwärts zu diesem CA- bzw. CG-Dinukleotid konnten weder DPE- noch MTE-Motive kartiert werden.

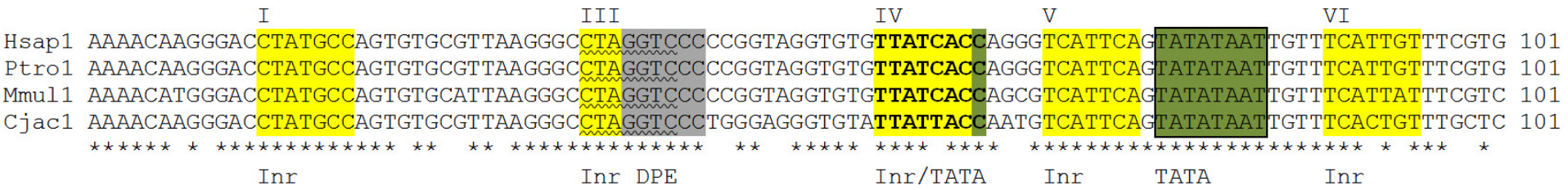


Abb. 4-20: Sequenzalignment der *MSY2* Repeateinheit 1 zwischen *Homo*, *Pan*, *Macaca* und *Callithrix*.

Die farbig markierten Blöcke und römischen Ziffern der Inr-Blöcke geben die Beschriftung aus Abb. 3 (*MSY2*-Vergleich Human) wieder. Farbcode: TATA-Motive sind grün, Inr-Motive sind gelb und DPEs sind hellgrau dargestellt. Das Inr-Motiv in Block III ist gewellt unterstrichen. Das TATA-Motiv in Block IV ist zur besseren Darstellung in Fettschrift dargestellt.

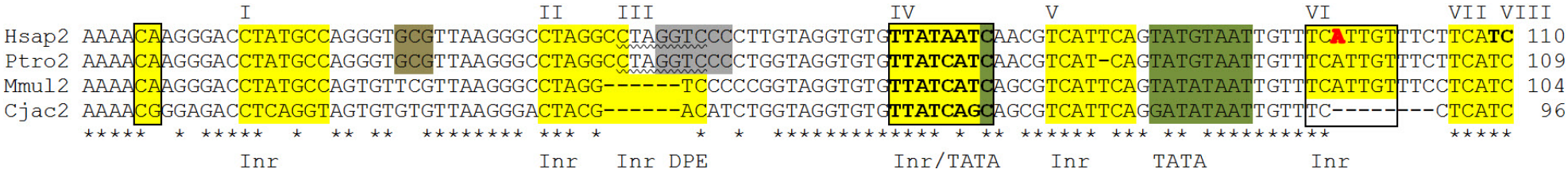


Abb. 4-21: Sequenzalignment der *MSY2* Repeateinheit 2 zwischen *Homo*, *Pan*, *Macaca* und *Callithrix*.

Die farbig markierten Blöcke und römischen Ziffern der Inr-Blöcke geben die Beschriftung aus Abb. 4-3 (*MSY2*-Vergleich Human) wieder. Die Human *DDX3Y* T-TSS-II (Hsap BAC-Position 54.051) ist in Inr-Block VI mit einem roten A markiert. Die positionshomologen Inr- und TATA-Motive der Human T-TSS-II sind durch Kästen in den vier Sequenzen markiert. Das Y-R Dinukleotid „CA/CG“ (Hsap BAC-Position 53.960), welches mit dem TATA-Motiv in *MSY2* Repeat 1 (zwischen Inr-Block V & VI) assoziiert scheint, ist gelb markiert. Das „GCG-Motiv“ in Position +20 stromabwärts des CA-Dinukleotides in Hsap und Ptro ist braun markiert. Die palindromische Sequenz „CCTAGG“ im Bereich von Inr-Block II und III ist nur in den Hominiden verdoppelt. Hierdurch entsteht ein zusätzliches Inr-Motiv (Inr-Block III; gewellt unterstrichen) und ein DPE (grau markiert). Im Bereich von Block VI ist die Sequenz „ATTGTTTC“ nur in *Callithrix* nicht verdoppelt. Hierdurch entfällt das Inr-Motiv in Inr-Block VI. Farbcode: TATA-Motive sind grün, Inr-Motive sind gelb, DPEs sind hellgrau und das GCG-Motiv ist orange dargestellt. Das TATA-Motiv in Block IV ist in Fettschrift dargestellt. Das *MSY2* Repeat 2/3 überlappende Human Inr-Motiv in Block VIII (in Fettschrift dargestellt) ist, auf Grund der fehlenden Vervierfachung des *MSY2* Repeats, nicht in anderen Primatenspezies vorhanden.

Auch das erst kürzlich definierte zweiteilige „Bridge Element“ ist nicht zu finden (Theisen et al., 2010). Für Mensch und Schimpanse konnte genomisch in der Position +20, stromabwärts des CA-Dinukleotides, das „GCG Motiv“ identifiziert werden. Das „GCG Echo-Motiv“ in Position +30 ist nicht vorhanden. Das „GCG Motiv“ wurde als Alternative zu MTE und DPE vorgeschlagen (Xi et al., 2007; Frith et al., 2008; Juven-Gershon et al., 2008). In den genomischen *Macaca* und *Callithrix* Sequenzen ist das „+20 GCG Motiv“ nicht vorhanden (Abb. 21).

Die Nutzung von *DDX3Y* T-TSS-I und T-TSS-II in nicht-human Primaten wird in den folgenden RT-PCR Analysen überprüft.

Pan troglodytes:

In Abb. 4-22 ist die RT-PCR Analyse von Schimpansen *DDX3Y* im Bereich des *MSY2* Repeats und Exon-T im Vergleich von Hodengewebe und lymphoblastoider Zellinie zu sehen. Es weist daraufhin, daß auch für *Pan* ein hodenspezifisches *DDX3Y* Exon-T Transkript existiert. Die Überprüfung in Schimpansen Nieren, Leber- und Milzgewebe bestätigte die Spezifität für Hodengewebe, da keine PCR-Amplifikate in diesen Geweben detektiert werden konnten (nicht gezeigt). Das negative PCR-Ergebnis mit P1328/2419 schließt auch für *P. troglodytes* eine Transkriptionsinitiation vor dem *MSY2* Repeat aus. Mit P1329/2419 sind zwei Amplifikate auf Grund der Primerbindung in Repeat 1 und 2 möglich. Es konnte aber nur das kürzere, für Repeat 2 spezifische, PCR-Produkt mit 580 bp Länge dargestellt werden (Abb. 4-22, Spalte 2).

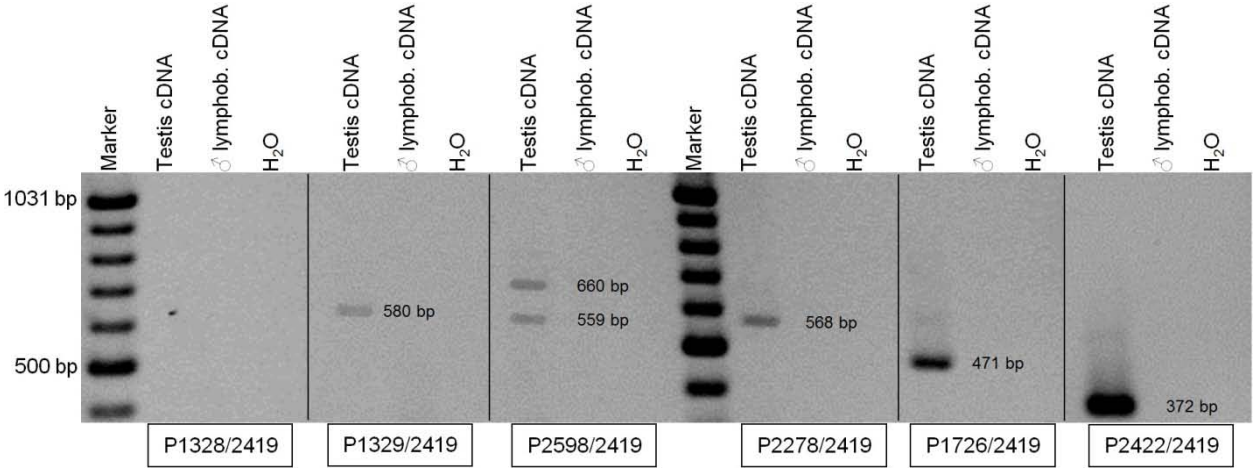


Abb. 4-22: RT-PCR Analyse *Pan troglodytes DDX3Y* Transkriptstart im Bereich *MSY2* und Exon-T. Errechnete PCR-Amplifikatgrößen: P1328/2419 = 736 bp, P1329/2419 = 681 u. 580 bp, P2598/2419 = 660 bp, P2278/2419 = 568 bp, P1726/2419 = 471 bp, P2422/2419 = 372 bp. Agarosegel ist im Bereich P1329 und P2598 schräg im elektr. Feld gelaufen. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Mit P2598/2419, der für Repeat 1 entwickelt wurde, konnten zwei sehr schwache Banden detektiert werden. Trotz Unterschiedes im ersten und letzten Nukleotid der P2598 Primer-sequenz band er ebenfalls an Repeat 2. Mögliche Inr-Motive bzw. Y-R-Dinukleotide sind in Abb. 4-15 markiert. Die Transkriptmengen, die mit P1329/2419 und P2598/2419 detektierbar sind,

waren allerdings so gering, daß eine funktionelle Bedeutung für die Schimpansen *DDX3Y* Genexpression fraglich erscheint.

Deutlich sichtbare Ptro *DDX3Y* PCR-Amplifikate im *MSY2* Repeat konnten erst mit P2278/2419 gefunden werden. Die P2278 Primerposition ist stromabwärts des oben beschriebenen Core-Promoter Moduls aus TATA-Box und CA-Dinukleotid lokalisiert. Dieses Modul ist folglich ein erster Kandidat für eine Schimpansen *DDX3Y* TSS im *MSY2* Minisatelliten-Repeat. Es fällt auf, daß mit P1726/2419 (vor T-TSS-I) und P2422/2419 (nach T-TSS-I) deutlich mehr PCR-Produkt amplifiziert wurde, als mit Primer, die im Bereich von *MSY2* lokalisiert liegen (Abb. 4-22). Für P2422/2419 läßt sich dies durch die TSS im Exon-T (T-TSS-I) erklären, die offensichtlich eine starke Transkription ermöglicht. Daß mit P1726/2419 mehr Transkripte amplifiziert werden als mit P2278/2419, erklärt sich durch die transkriptionelle Aktivität mehrerer Initiator-Motive im *MSY2* Repeat 2, wodurch eine graduelle Verstärkung der Transkription erfolgt. So befindet sich zwischen den P2278 und P1726 Primerlokalisationen auch das Core-Promoter Modul (TATA-Box & Inr) welches in Human als *DDX3Y* T-TSS-II definiert wurde (Abb. 4-2 & 4-15).

Die RT-PCR Ergebnisse weisen darauf hin, daß in der Schimpansen *MSY2* und Exon-T Region ebenfalls zwei Initiationsstellen T-TSS-I und T-TSS-II vorhanden sind. Die Hauptmenge an *P. troglodytes DDX3Y* Exon-T Transkripten wird offensichtlich von T-TSS-I aus transkribiert. Eine Sequenzanalyse verifizierte die PCR-Amplifikate als *DDX3Y* Exon-T Transkripte. Zudem wurde durch die Sequenzanalyse bestätigt, daß ausschließlich *DDX3Y* Exon-Ta-1b Transkripte in Schimpansen Hodengewebe synthetisiert werden. Weitere Exon-T Splicevarianten konnten nicht dargestellt werden.

Macaca mulatta:

In den Abb. 4-23 u. 4-24 sind die Ergebnisse der *Macaca mulatta DDX3Y* RT-PCR Analyse zu sehen. Die *DDX3Y* Transkriptinitiation im *MSY2* Repeat wird durch die PCR-Amplifikate mit den Primerpaaren P1329/1286 und P2531/1286 auch für *Macaca* bestätigt (Abb. 4-23). Die Analyse in weiteren *Macaca*-Geweben (Niere, Leber, Milz) zeigte keine signifikante Transkriptionsaktivität (Rauschendorf et al., 2011; hier nicht gezeigt). Mit dem in *MSY2* Repeat 1 gelegenen Primer P2599 konnte kein PCR-Amplifikat erhalten werden (Abb. 4-23). Allerdings waren in der negativen Kontrolle P2601/1286 (vor *MSY2* lokalisiert) stets sehr schwache PCR-Amplifikat der errechneten Größe zu sehen (Abb. 4-23). Eine Klonierung und anschließende Sequenzierung des P2601/1286 PCR-Produktes war allerdings nicht möglich, so daß eine Initiation 5´ des *Macaca MSY2* Repeats nicht ausgeschlossen werden kann. Eine erste potentielle *DDX3Y* TSS in *MSY2* sollte dennoch zwischen den P2599 und P1329 Primerpositionen zu finden sein, da mit P2601 keine starken PCR-Amplifikate zu erhalten waren, die eindeutig für ein funktionelles

Transkript sprechen würden. Als mögliches Core-Motive für die *M. mulatta DDX3Y* Transkriptinitiation in *MSY2* ist, wie bei *P. troglodytes*, das Modul aus TATA und dem CA-Dinukleotid am 5´ Ende von Repeat 2 zu benennen, wofür auch das nochmals stärkere PCR-Amplifikat mit P2531/1286 spricht (Abb. 4-23). Der Forward-Primer P2531 ist unmittelbar stromabwärts des CA-Dinukleotides lokalisiert. Mögliche Inr-Motive bzw. Y-R-Dinukleotide sind in Abb. 4-15 markiert. Im Unterschied zu *P. troglodytes* ist bei *M. mulatta* bereits innerhalb des *MSY2* Repeats eine starke *DDX3Y* Transkription zu beobachten.

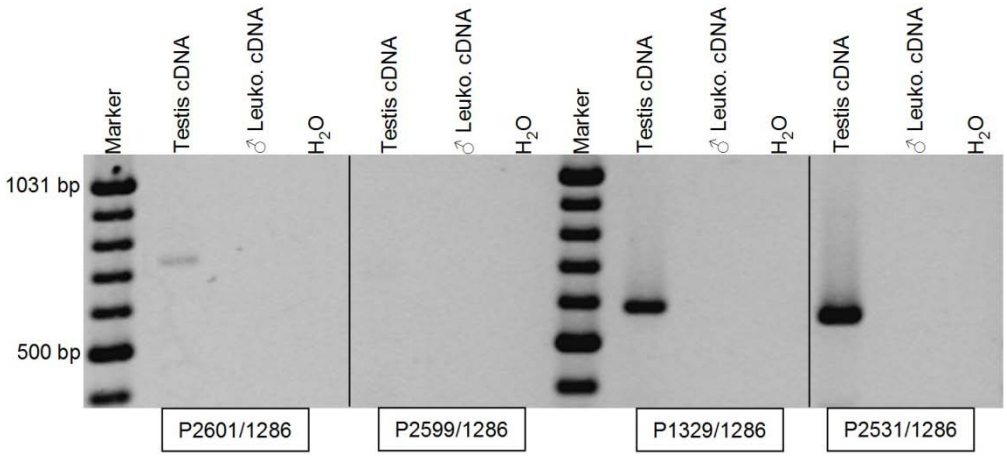


Abb. 4-23: RT-PCR Analyse *Macaca mulatta DDX3Y* Transkriptstart im Bereich *MSY2*.
Errechnete PCR-Amplifikatgrößen: P2601/1286 = 738 bp, P2599/1286 = 672 bp, P1329/1286 = 577 bp, P2531/1286 = 565 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100 bp DNA-Leiter.

In Abb. 4-24 ist zu erkennen, daß im Vergleich der zwei *M. mulatta DDX3Y* RT-PCRs mit den Forward-Primern P1726 und P2530 nochmals eine Steigerung der PCR-Amplifikatmenge vorliegt, was an der Lokalisation der *DDX3Y* T-TSS-I zwischen diesen Primerpositionen liegen dürfte. Auch stromabwärts der *DDX3Y* T-TSS-I sind für *Macaca* die Transkripte nur hodenspezifisch festzustellen. Diese *Macaca* PCR-Produkte konnten ebenfalls per Sequenzanalyse als *DDX3Y* Exon-Ta-1b Transkripte bestätigt werden. Wie auch für P. troglodytes, konnten von *M. mulatta* keine weiteren Splicevarianten von *DDX3Y* Exon-T identifiziert werden. RT-PCR Analysen mit dem Exon-Tb spezifischen Primer P1727 (Lokalisation siehe Abb. 4-15) waren sowohl für *Pan* als auch für *Macaca* negativ (nicht gezeigt). Ausschließlich in Human Hodengewebe konnte mit P1727 die *DDX3Y* Exon-Tb Transkriptvariante dargestellt werden und stellt sich somit als humanspezifisches *DDX3Y* Splicing dar (Zimmer J., unveröffentlichte Resultate).

Die *Macaca mulatta DDX3Y* Exon-T Transkription erfolgt, wie bei den Hominiden Mensch und Schimpanse, ebenfalls exklusiv in Hodengewebe und initiiert wahrscheinlich ebenso in den zwei bekannten Positionen T-TSS-I in Exon-T und T-TSS-II in *MSY2* Repeat 2. Eine graduelle Steigerung der Transkriptmenge, durch Nutzung multipler Core-Motive innerhalb des *MSY2* Repeats, ist wahrscheinlich.

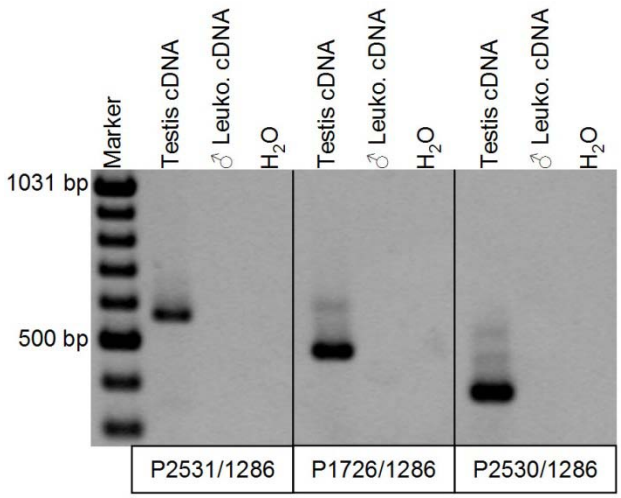


Abb. 4-24: *Macaca mulatta* RT-PCR im Bereich von *DDX3Y* Exon-T.
Errechnete PCR-Amplifikatgrößen: P2531/1286 = 565 bp, P1726/1286 = 473 bp, P2530/1286 = 374 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Callithrix jacchus:

Die RT-PCR Analyse in *Callithrix jacchus* ergab keine *DDX3Y* Transkriptinitiation im Bereich von *MSY2* (Abb. 4-25). Dies ist ein erster großer Unterschied, der den Neuweltaffen *Callithrix* klar von den Catarrhini (Hsap, Ptro, Mmul) abgrenzt.

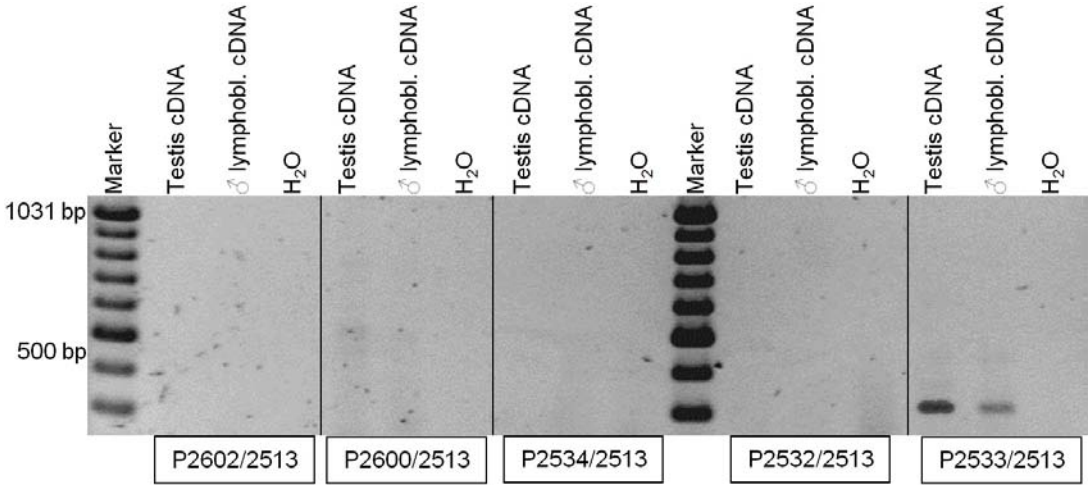


Abb. 4-25: *Callithrix jacchus DDX3Y* RT-PCR im Bereich von *MSY2* und Exon-T.
Errechnete PCR-Amplifikatgrößen: P2602/2513 = 666 bp, P2600/2513 = 586 bp, P2534/2513 = 498 bp, P2532/2513 = 414 bp, P2533/2513 = 315 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100 bp DNA-Leiter.

Mit den Primerpaaren P2602/2513 (negativ Kontrolle vor *MSY2*), P2600/2513 (*MSY2* Repeat 1), P2534/2513 (*MSY2* Repeat 2) und P2532/2513 (nach *MSY2*) konnten keine *DDX3Y* PCR-Produkte dargestellt werden. *Callithrix DDX3Y* PCR-Produkte sind erst mit P2533/2513 amplifizierbar (Abb. 4-25). Der Forward-Primer P2533 (homolog zu Hsap P1313) ist stromabwärts der *DDX3Y* T-TSS-I Position lokalisiert (Abb. 4-15).

Ein weiterer Unterschied zwischen *Callithrix* und den Catarrhini ist die Tatsache, daß für *Callithrix DDX3Y* Exon-T keine vollständige Hodenspezifität gezeigt werden konnte. Mit P2533/2513 wurden PCR-Amplifikate in cDNAs aus Hodengewebe und der lymphoblastoiden

Zellinie detektiert (Abb. 4-25). Insgesamt konnte eine präferentielle Transkription in Hodengewebe festgestellt werden. Es waren in geringerem Maße auch PCR-Produkte in Hirn-, Nieren- und Lebergewebe, sowie in der lymphoblastoiden Zellinie zu detektieren (hier nicht gezeigt). Die Sequenzanalyse bestätigte auch für *Callithrix jacchus* diese PCR-Produkte als *DDX3Y* Exon-Ta-1b Transkripte.
Um zu evaluieren, ob neben dem qualitativen Unterschied, wie der Anzahl der *MSY2* Repeateinheiten, auch quantitative Unterschiede durch die Vervierfachung des *MSY2* Repeats in *H. sapiens* bestehen, wurde eine semi-quantitative RT-PCR mit Hodengewebe der vier Primatenspezies durchgeführt (Abb. 4-26).

In Reihe A der Abb. 4-26 wird deutlich, daß mit dem in Repeateinheit 2 lokalisierten Primerset (Human P2278), nur in Mensch, Schimpanse und *Macaca DDX3Y* Transkripte gezeigt werden können. Die RT-PCR mit 30 Zyklen zeigt zudem, daß die *DDX3Y* Expressionsrate in Schimpanse geringer, als in Hodengewebe von Mensch und Rhesusaffe, ist.
In Reihe B wird das Primerset (*Human* P1726) direkt stromabwärts von *MSY2* und 5´ der T-TSS-I Position verwendet. Auch hier ist deutlich zu sehen, daß nur in den Catarrhini *DDX3Y* Transkripte vorhanden sind, die in *MSY2* starten. Erst mit dem dritten Primerset (Human P1313) stromabwärts von T-TSS-I ist auch eine Expression in *Callithrix* Hodengewebe darstellbar (Reihe C in Abb. 4-26). Für den Schimpansen fällt auf, daß im Vergleich zu Mensch und *Macaca* die T-TSS-I Position in Exon-T wesentlich mehr genutzt wird als die T-TSS-II Position im *MSY2* Repeat. In Abb. 4-26 ist weiterhin auffällig, daß das *Callithrix DDX3Y* Exon-T Transkript kürzer ist und schwächer exprimiert wird, als die korrespondierenden *DDX3Y* Transkripte in den Catarrhini.

In diesem Kapitel konnte somit gezeigt werden, daß das *DDX3Y* Exon-T nicht exklusiv in Human exprimiert wird, sondern daß dieses eine gemeinsame Eigenschaft der Primaten darstellt. Die *DDX3Y* T-TSS-I Position wird von allen vier untersuchten Primatenspezies genutzt. In den Catarrhini sind die in T-TSS-I startenden Transkripte hodenspezifisch exprimiert, wohingegen in *Callithrix* nur eine präferentielle, aber nicht exklusive Transkription in Hodengewebe festgestellt werden konnte. Nur den Catarrhini gemein ist ein zweiter hodenspezifischer Transkriptstart in dem *MSY2* Minisatelliten-Repeat, der in dem Neuweltaffen *Callithrix jacchus*, auf Grund von Sequenzunterschieden in der *MSY2* Repeateinheit 2, nicht realisiert ist.

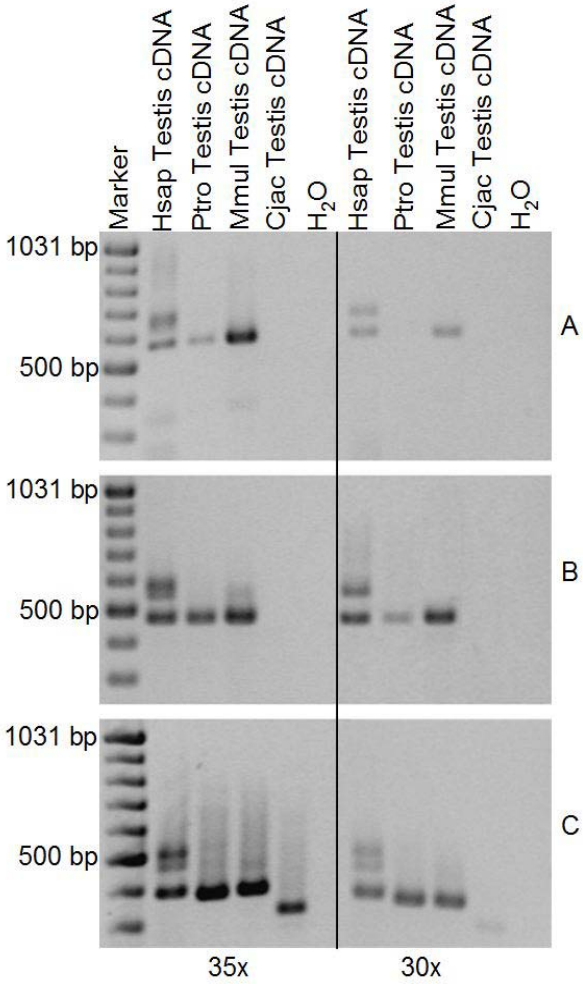


Abb. 4-26: Vergleich der *DDX3Y* Expression im Bereich von *MSY2* und Exon-T im Hoden-gewebe von Mensch, Schimpanse, Rhesusaffe und Weißbüschelaffe.

Der Vergleich erfolgt mittels semi-quantitativer RT-PCR mit 35 und 30 PCR-Zyklen.

A: Verwendete Primer sind für Mensch P2278/1286, für Schimpanse P2278/2419, für *Macaca* P2531/1286 und für *Callithrix* P2534/2513.

B: Verwendete Primer sind für Mensch P1726/1286, für Schimpanse P1726/2419, für *Macaca* P1726/1286 und für *Callithrix* 2532/2513.

C: Verwendete Primer sind für Mensch P1313/1286, für Schimpanse P2422/2419, für *Macaca* P2530/1286 und für *Callithrix* 2533/2513. Die *C. jacchus DDX3Y* Sequenz ist auf Grund von indels um circa 100 bp kürzer. DNA-Marker ist eine 100 bp DNA-Leiter.

Ein Unterschied in der *DDX3Y* Transkriptinitiation in *MSY2* der Catarrhini scheint eher in der Transkriptionsstärke als in verschiedenen zur Initiation genutzten Core-Promoter Motiven zu bestehen. Die Schimpansen und *Macaca* T-TSS-II (TATA & CA-Dinukleotid) konnte in den 5´ Bereich des zweiten Repeats kartiert werden. Der Startpunkt der Human *DDX3Y* T-TSS-II liegt zwar am 3´ Ende des zweiten Repeats, könnte aber durch die Vervierfachung des *MSY2* von dem ursprünglichen TATA & CA-Dinukleotid dorthin verschoben worden sein.

Einen Unterschied zwischen Mensch und nicht-Human Primaten stellen die *DDX3Y* Exon-T Splicevarianten dar. Die RT-PCR- und Sequenzanalysen bei *Pan*, *Macaca* und *Callithrix* wiesen nur das Exon-Ta-1b Splicemuster auf, obwohl die notwendigen Signale für die Splicevarianten Ta-1a und Tb-1b bei den Catarrhini alle vorhanden sind. Eine homologe Splicevariante Ta-1a ist bei *Callithrix* nicht möglich, da das Exon-1a AG-Spliceakzeptorsignal deletiert ist (Abb. 4-15).

4.4.2. Evolution der *DDX3Y/Ddx3y* Keimbahn Core-Promoterstruktur in Rind und Maus

Der Vergleich der Human *DDX3Y* Core-Promoter Regionen mit zwei nicht-Primaten Säugern, Rind und Maus, soll darüber Aufschluß geben, ob die beobachteten *DDX3Y* 5´ Transkriptverlängerungen eine exklusive Eigenschaft der Primaten darstellen, oder ob solche Transkriptvarianten in gleichem Maße in Vertretern anderer Säugerspezies vorhanden sind.

Zunächst wurden per BLAST-Suche korrespondierende genomische Sequenzen von Rind (*Bos taurus* = Btau) und Maus (*Mus musculus* = Mmus) analysiert, in denen vollständige *Ddx3y* Sequenzen zu identifizieren waren. Für das Rind konnte der Y-chromosomale BAC „CH240-56A13“[26] mit der *Ddx3y* Genregion von Position 140.763 bis 127.423 (*MSY2* bis Ende Exon 17) identifiziert werden. Für die Maus konnte der Y-chromosomale BAC „RP24-208N6“[27] mit der *Ddx3y* Genregion von Position 94.372 bis 68.816 (*MSY2* bis Ende Exon17) gefunden werden. Beide zu analysierende *Ddx3y* Gene verfügen über dieselbe Exonanzahl, wie das Human *DDX3Y* Transkript NM_004660.3, über ein Translationsstart- und Stopcodon sowie den Polyadenylierungsstellen. Die Homologien in den kodierenden *Ddx3y* Sequenzen betragen 88% für die Maus und 89% für das Rind. Die Tabellen mit den Rind und Maus *Ddx3y* Exon-Intronstrukturen befinden sich im Anhang (Tab. A-8, A-9).
Mit Hilfe der BLAST und ClustalW Sequenzvergleiche konnten auch in den *Ddx3y* 5´ Regionen von Rind und Maus potentiell *MSY2*-homologe Sequenzabschnitte identifiziert werden. Für Rind und Maus konnte jeweils eine *MSY2*-Sequenz identifiziert werden, die mit der Human *MSY2-1* Sequenz homologisierbar ist (Abb. 4-27). Das Sequenzalignment in Abb. 4-27 zeigt vor allem zwischen Mensch und Maus eine erhebliche Sequenzdivergenz. Dies findet Ausdruck in einer 421 bp kürzeren Maus-Sequenz im Vergleich der *DDX3Y/Ddx3y* Regionen. Im Vergleich zu Human sind es vor allem Sequenzabschnitte in Exon-T und Intron-T, die in der Maussequenz nicht vorhanden sind. Das dargestellte Sequenzalignment ist manuell aus mehreren einzelnen ClustalW-Alignments zusammengefügt, da ClustalW kein adäquates Gesamt-Sequenz-alignment über die angegebene Distanz, auf Grund der erheblichen Sequenzdivergenzen darstellen kann. Für das Alignment dienten die identifizierten *MSY2* Regionen und die kodierenden *Ddx3y* Exon1 Bereiche als Fixpunkte (Abb. 4-27). Aus dem Sequenzalignment wird deutlich, daß die Sequenzunterschiede trotz der größeren evolutionären Distanz zwischen Mensch und Rind geringer ausfallen, als zwischen Mensch und Maus (Abb. 4-27). Für die kodierenden *DDX3Y/Ddx3y* Bereiche konnte, im Gegensatz zu den regulativen 5´ Bereichen, ein hoher Grad an Homologie von ~ 90% in allen drei Spezies festgestellt werden.

---

[26] Btau: GenBank Acc.No. AC217167.7; revers-komplementär.
[27] Mmus: GenBank Acc.No. AC145393.4; revers-komplementär.

Abb. 4-27: Genomisches *DDX3Y/ Ddx3y* Sequenzalignment zw. Mensch (Hsap), Rind (Btau) und Maus (Mmus).

Human-Sequenz von BAC-Position 53.700 bis 55.190 (475I1; vor *MSY2* bis Ende Exon 1). Rind-Sequenz von BAC-Position 140.791 bis 139.747 (CH240-56A13). Rind *Ddx3y* ATG ist in Position 139.791. Maus-Sequenz von BAC-Position 94.461 bis 93.700 (RP24-208N6). Maus *Ddx3y* ATG ist in BAC-Position 93.744. Kartierte Human Core-Promoter Motive in der *DDX3Y* Sequenz sind entsprechend Abb. 4-3 markiert. Inr-Motive sind gelb markiert und TSSs mit roten Buchstaben A und G hervorgehoben. TATA-Box Motive sind grün, MTE- hellblau, DPE-Motive sind hellgrau dargestellt. Das Human *DDX3Y* TSS-Id Inr-Motiv fällt mit einem DPE zusammen, welches gewellt unterstrichen dargestellt ist. Das Start-codon ATG ist rot hervorgehoben, die 5´ und 3´Splice-Signale sind in Fettschrift und pink angegeben. Ermittelte Splice-signale in den Rind *Ddx3y* Transkripten sind ebenfalls eingezeichnet. Die pot. positionshomologen Sequenzen zu den Human Core-Promoter Elementen sind gerahmt dargestellt. Potentielle Core-Promoter-Motive in Rind und Maus Sequenz sind orange eingezeichnet. Der Y-spezifische *MSY2* Repeat ist durch vertikale Striche gekennzeichnet. In der Human Sequenz sind die *MSY2* Repeateinheiten 2-4 nicht angegeben und durch zwei vertikale Striche gekennzeichnet. Sterne unterhalb des Alignments geben homologe Nukleotide in allen drei Sequenzen an. Punkte unterhalb des Alignments markieren konservierte Nukleotide in den Human und Rind Sequenzen, Doppelpunkte geben konservierte Nukleotide zwischen Human und Maus und Pluszeichen markieren konservierte Nukleotide zwischen Rind und Maus. Der Maus *Ddx3y* Primer P2801 ist im Alignment nicht gezeigt und befindet sich stromaufwärts der gezeigten Sequenz.

Maus *Ddx3y* Core-Promoter & RT-PCR Analyse:

Für Maus *Ddx3y* konnte eine erste TSS-Kartierung, mit den dazugehörigen Gewebeinformationen, aus der DBTSS ermittelt werden (Tab. 4-5). Von 11 kartierten Einträgen aus der DBTSS enthalten nur fünf das Y-R bzw. Inr-Motiv. Lediglich ein cDNA-Klon mit Start in Position 93.758 entspricht der Inr-Konsensussequenz. Der potentielle Maus TSS-Bereich aus der DBTSS beträgt 65 bp (Pos. 93.822 - 93.758) und entspricht somit dem Typus eines breiten Core-Promoters.

Tab. 4-5: Positionen potentieller Initiator-Motive im Bereich des ubiquitären Maus *Ddx3y* Transkriptstartes in Exon1 (TSS-I Region).

Mismatches der ermittelten Sequenzen zur Inr-Konsensussequenz sind rot markiert. Das hochkonservierte Y-R Dinukleotid ist in Spalte 2 unterstrichen. Initiator-Motive mit 100% Übereinstimmung zur Konsensussequenz sind im Fettdruck und grün hervorgehoben. Weitere potentielle TSS mit assoziiertem Y-R Dinukleotid sind hellgrau dargestellt. TSSs die nicht der Inr-Seq. entsprechen bzw. kein Y-R Dinukleotid enthalten, sind farblich nicht unterlegt. Die Positionsangaben entsprechen der +1 Position (= TSS) des Transkriptes, die wiederum dem dritten Nukleotid in der Inr-Sequenz entspricht. In Spalte 4 wird angegeben aus welchen Geweben/Zellen die Transkripte isoliert wurden. Die Zahlenangaben in Spalte 5 geben die Anzahl der Nennungen in der DBTSS an. Positionsangaben beziehen sich auf den murinen Y-BAC RP-208N6. Maus *Ddx3y* ATG ist an Position 93.744.

Position der 5´Enden der DBTSS Klone	Sequenz des pot. Inr Motives: YYANWYY	Y-R Regel für -1/+1 Position im Inr	Gewebe und Anzahl cDNA-Klone pro Gewebe	Gesamtzahl cDNA-Klone in DBTSS
93.850	CAGTACG	-	Plazenta	1
93.822	GTACGTA	+	Total-Embryo	1
93.821	TACGTAA	-	Hirn	1
93.820	ACGTAAG	+	Herz, Hirn	2
93.801	CCGTGAG	+	Lunge, Hirn	2
93.796	AGGAAAC	-	Innenohr (2)	2
93.795	GGAAACA	-	Herz	1
93.793	AAACAGG	-	Plazenta	1
93.792	AACAGGT	-	Embryo-Kopf, Rückenmark, Herz	3
93.791	ACAGGTC	+	Blastozysten, dendritische Zellen (2), embryonale Rathke-Tasche, Augapfel	5
93.758	**TCAATTT**	+	Hoden	1

In der Region zwischen der Position der Human TSS-I-(T) und TSS-I konnten in der Maus *Ddx3y* Sequenz mindestens neun potentielle TATA-Motive identifiziert werden, die teilweise mit den DBTSS-Positionen assoziiert scheinen. Dies ist z.B. der Fall für den cDNA-Klon mit Start in Position 93.791 (Tab. 4-5). Mit fünf cDNA-Klonen aus vier unterschiedlichen Mausgeweben, ist dies die einzige Position, die häufiger als Startpunkt genannt wird. An Position -32 zu dieser TSS befindet sich ein pot. TATA-Motiv „TACGTAAG“ (ein Mismatch) positionshomolog zu dem TATA-Motiv der Human *DDX3Y* TSS-I Region. Stromabwärts des Inr-Motives „ACAGGTC“ in Position 93.791 konnte in Position +28 ein DPE-Motiv „GTTCTC“ mit zwei Mismatches kartiert werden.

Die TSS in Position 93.758, mit dem kanonischen Inr-Motiv „TCAATTT“, ist 14 Nukleotide stromaufwärts des *Ddx3y* ATG-Startcodons lokalisiert. Stromaufwärts und -abwärts konnten zu dieser TSS weder TATA-Motive, noch MTE oder DPE-Motive kartiert werden.
Aus den NCBI- und Ensembl-Datenbanken konnten zwei weitere potentielle TSS-Positionen ermittelt werden. Das Maus *Ddx3y* Transkript aus NCBI (GenBank Acc.No. NM_012008.2) startet in Position 93.834 und das Transkript (ENSMUST00000091190) aus der Ensembl-Datenbank initiiert in Position 93.798. Diese beiden Transkripte sind 5´ nicht vollständig. Der NCBI cDNA-Klon in Position 93.834 liegt zwischen den P2328 und P2313 Primerpositionen (Abb. 4-27) und markiert laut DBTSS wahrscheinlich ungefähr das Ende der ubiquitären *Ddx3y* TSS-Region.

Zur experimentellen Überprüfung der Maus *Ddx3y* TSS-Region wurden gezielte RT-PCR Experimente, in Maus Hodengewebe und in männlichen Leukozyten, durchgeführt (Abb. 4-28). Die Primerpositionen sind der Abb. 4-27 zu entnehmen.

Mit allen Primerpaaren konnten quantitative Unterschiede in der Expression zwischen Hodengewebe und Leukozyten festgestellt werden. Mit P2328/2311 (Pos. 93.828) ist dieser Unterschied in der Transkriptmenge am deutlichsten feststellbar (Abb. 4-28). Mit dem 36 Nukleotide weiter stromaufwärts gelegenen Primer P2371, sind ausschließlich Transkripte im Hodengewebe der Maus darstellbar. Mit dem Primerpaar P2372/2311 sind schließlich auch in Hodengewebe keine PCR-Amplifikate mehr zu detektieren.

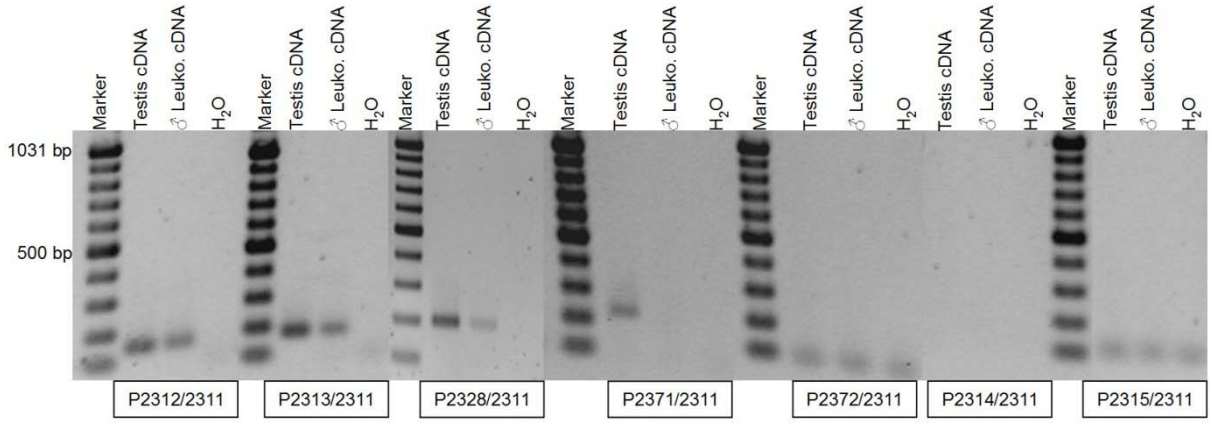


Abb. 4-28: *Ddx3y* RT-PCR Analyse *Mus musculus*.
Errechnete PCR-Amplifikatgrößen: P2312/2311 = 142 bp, P2313/2311 = 166 bp, P2328/2311 = 195 bp, P2371/2311 = 224 bp, P2372/2311 = 274 bp, P2314/2311 = 496 bp, P2315/2311 = 579 bp. 35 PCR-Zyklen. DNA-Marker ist 100bp DNA-Leiter.

Mit den Forwardprimer P2314 und P2315, aus der positionshomologen Maus Exon-T-Sequenzen, konnten keine Maus *Ddx3y* Transkripte amplifiziert werden, so daß ein Start in Maus Exon-T ausgeschlossen wird. Die von Human *DDX3Y* bekannten Splicesignale Exon-Ta,

Exon-Tb und Exon-1a sind genomisch in Maus nicht konserviert. Mit dem in *MSY2* lokalisierten Forward-Primer P2802 und dem stromaufwärts von *MSY2* gelegenen Primer P2801 konnten ebenfalls keine Maus *Ddx3y* Amplifikate dargestellt werden. Diese RT-PCRs werden in Abb. 4-28 nicht gezeigt, da bereits P2372, P2314 u. P2315 keine Amplifikate mehr gezeigt haben.

Die ubiquitäre Maus *Ddx3y* Core-Promoter Region ist somit höchstwahrscheinlich zwischen den Primerpositionen P2312 und P2328 lokalisiert. Dieses Ergebnis stimmt mit den DBTSS-Daten aus Tabelle 4-5 überein. Als Hauptkandidat für die ubiquitäre Maus *Ddx3y* Transkription kommt somit die oben beschriebene TSS in Position 93.791 in Frage („ACAGGTC“; Nr. 10, Tab. 4-5). Die Maus *Ddx3y* Transkripte der ubiquitären TSS-I Region haben somit eine 5´UTR Länge von maximal circa 90 Nukleotiden.

Zusätzlich konnte mit P2371 gezeigt werden, daß auch für die Maus ein zweiter aktiver Core-Promoter in der TSS-I Region existiert, der nur in Hodengewebe genutzt wird (TSS-I-T). Dieser gewebespezifisch genutzte Core-Promoter ist stromaufwärts von P2371 zu suchen.
Ein korrespondierendes Core-Promotermotive wäre 5´ der P2372 Primerposition das kanonische TATA-Box Motiv der Sequenz „TTAAATAG“ (Abb. 4-27). Das dazugehörige Inr-Motiv „TTGACGA“ bzw. Y-R-Dinukleotid ist in Position 93.893 (für +1 TSS) zu finden. Das potentielle TATA-Motiv wäre in der relativen Position -29 zu dieser potentiellen TSS. Stromabwärts ist im korrekten Abstand von +28 zusätzlich ein kanonisches DPE-Motiv „TGACGC“ zu kartieren. Das potentielle Core-Promoter-Modul für einen Transkriptstart zwischen den Primerpositionen P2371 und P2372 ist in Abb. 27 markiert. Die längste, hodenspezifische 5´UTR der Maus *Ddx3y* Transkripte hätte damit eine Länge von 149 Nukleotiden.
Im Vergleich zur Human *DDX3Y* TSS-I und TSS-I-(T) Region ist zu vermerken, daß für Maus *Ddx3y* ebenfalls ein variabler Bereich für die ubiquitäre Transkription kartiert werden konnte. Der Maus Bereich zeigt sich mit maximalen 5´UTR Längen von 90 Nukleotiden dabei etwas kürzer als die entsprechende Human TSS-I Region mit max. 114 Nukleotiden. Ebenso konnte eine hodenspezifische TSS-I-(T) Position kartiert werden. Diese zeigt in der 5´UTR Länge einen deutlichen Unterschied. Die Human TSS-I-(T) 5´UTR ist mit 235 Nukleotiden fast 90 nts länger als die entsprechende Maus-Sequenz. Die Analyse hat auch ergeben, daß wahrscheinlich nicht positionshomologe Core-Promoter Motive genutzt werden, wie zwischen Mensch und den anderen Primaten, sondern daß Mensch und Maus jeweils eigene Core-Promoter Domänen in den *DDX3Y*/*Ddx3y* TSS-I Regionen verwenden, um in beiden Spezies eine ubiquitäre TSS-I und eine hodenspezifische TSS-I-(T) Region zu etablieren.

Rind *Ddx3y* Core-Promoter & RT-PCR Analyse:

Da in der DBTSS Datensammlung keine Daten zu Rind *Ddx3y* verfügbar waren, diente ein Rind *Ddx3y* Transkript (GenBank Acc.No. NM_001172595.1)[28] aus der NCBI-Datenbank, mit Start in BAC-Position 139.910 (BAC CH240-56A13), als Startpunkt für die experimentelle Rind TSS-Analyse. Die BAC-Position 139.910 ist zwischen Primer P2570 und P2569 lokalisiert (Abb. 4-27). Diese cDNA entspricht nicht der Y-R-Dinukleotidregel („GGTCATT“) und könnte somit 5´ unvollständig sein. Allerdings befindet sich in der relativen Position -28 ein kanonisches TATA-Box Motiv „TTTAAAGG“.

Wie auch in der Maus *Ddx3y* TSS-I Region fehlen beim Rind die homologen Sequenzen zu den kartierten Human Inr-Motiven TSS-Ia und -Id (Abb. 4-27). Eine variable Initiation der Ddx3y Transkription ist auch für das Rind möglich, da dennoch mehrere potentielle Core-Promotermotive in dieser TSS-I Region vorhanden sind.

Die Analyse erfolgte per RT-PCR, mit schrittweise nach 5´ verlagerten Oligonukleotiden. In Abb. 29 ist zu erkennen, daß *Ddx3y* bei Rind scheinbar gewebespezifisch genutzt wird. Mit allen getesteten Primer konnten RT-PCR Amplifikate ausschließlich in Hodengewebe detektiert werden. P2568/2567 diente als Kontrolle, ob mit einem Primer stromabwärts zum Rind *Ddx3y* Translationsstartcodon ATG, dennoch RT-PCR Amplifikate in anderen Geweben darstellbar sind. Abb. 4-29 zeigt, daß mit P2568/2567 keine signifikanten Amplifikatmengen in anderen Geweben erhältlich sind.

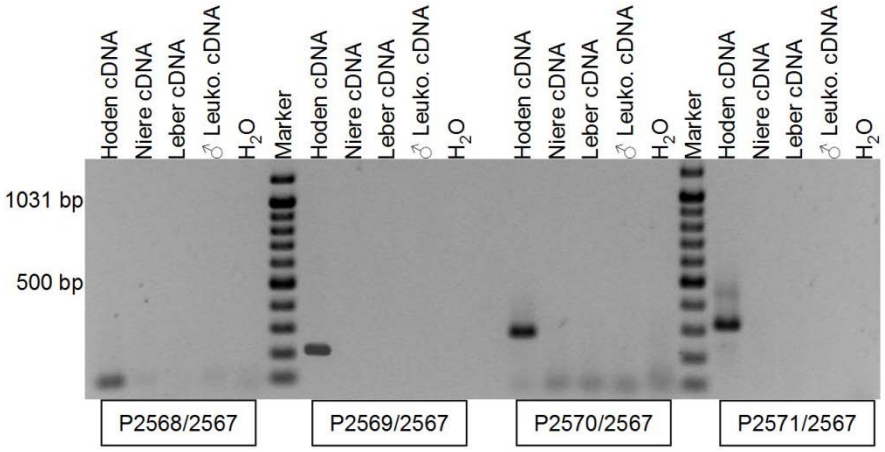


Abb. 4-29: RT-PCR Analyse Rind *Ddx3y*. Errechnete RT-PCR Amplifikatgrößen für ungesplicte Transkripte: P2568/2567 = 105 bp, P2569/2567 = 210 bp, P2570/2567 = 277 bp, P2571/2567 = 717 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Mit dem Forwardprimer P2570, der 48 Nukleotide vor dem in NCBI genannten Transkriptbeginn in BAC-Position 139.910 lokalisiert ist, konnte ein eindeutiges RT-PCR Amplifikat detektiert werden (Abb. 4-29). Stromaufwärts zu P2570 konnte positionshomolog zum Human Inr-Motiv in TSS-I-(T) die Sequenz „CTAGCTA“, mit zwei Mismatches zur Konsensussequenz und einem Y-R Dinukleotid an der -1/+1 Position, lokalisiert werden. An Position +28 ist ein potentielles DPE-

[28] Das Rind *Ddx3y* Transkript (GenBank Acc.No. NM_001172595.1) entspricht der cDNA-Sequenz aus einer Rinderhoden cDNA-Bibliothek [b*DDX3Y*-S (GenBank Acc.No. GQ259591)] in Liu et al. (2009).

Motiv „GGTAGA“ positionshomolog zum Human DPE-Motiv stromabwärts von TSS-I-(T) kartierbar. Als Startposition für das **„Rind *Ddx3y* TSS-I-(T)“** wird die Position 139.971 im Y-BAC CH240-56A13 vorgeschlagen. Mit einer 5´UTR-Länge von 180 Nukleotiden, ist das Rind *Ddx3y* Transkript 55 Nukleotide kürzer als das Human *DDX3Y* TSS-I-(T) Transkript.

Mit dem Forwardprimer P2571, der im homologen Exon-T Sequenzbereich lokalisiert ist, konnten ebenfalls RT-PCR Produkte amplifiziert werden (Abb. 4-29). Es ist zu sehen, daß das erhaltene RT-PCR Amplifikat nicht über die errechnete Länge von 717 bp, für ein durchgehendes Transkript, verfügt. Analog zum Human *DDX3Y* Transkript mit Exon-Ta-1b Splicing, sollten Rind *Ddx3y* RT-PCR Produkte eine Länge von 229 bp aufweisen. Das sichtbare Rind *Ddx3y* Amplifikat in Abb. 4-29 ist allerdings über 300 bp groß. Die Sequenzierung dieses RT-PCR Produktes ergab, daß es sich um ein Rind *Ddx3y* Transkript handelt, welches die in Mensch bekannte 5´ Exon-Ta (GT-) Donor-Splicestelle nutzt. Als 3´AG-Spliceakzeptorstelle wird eine, 21 bp stromaufwärts der bekannten Human Exon-1a gelegene, Rinder-spezifische 3´AG-Splicestelle genutzt (siehe Abb. 4-27). Zur Unterscheidung, zu den Human-Splicevarianten, wird sie **„Exon-1c“** genannt. Das erhaltene Rind *Ddx3y* RT-PCR Produkt hat eine Länge von 319 Nukleotiden (P2571/2567) und entsteht durch Exon-Ta-1c Splicing (Abb. 4-31).

Mit den weiteren Primer P2589, P2590, 2591 und P2603 soll die 5´Grenze dieses Rind *Ddx3y* Exon-T´s evaluiert werden (Abb. 4-30). Der unmittelbar stromabwärts zu *MSY2* positionierte Forwardprimer P2590 ergab ein PCR-Amplifikat, welches ein *Ddx3y* Transkript mit Exon-Ta-1c Splicing und einer PCR-Produktlänge von 536 Nukleotiden darstellt (Abb. 4-30). Auf Grund der Primerlage, stromaufwärts der positionshomologen Human *DDX3Y* T-TSS-I, ist auch für Rind ein Transkriptstart innerhalb der *MSY2*-Sequenz anzunehmen (Abb. 4-27).
Mit dem weiter stromabwärts gelegenen Forwardprimer P2589 (positionshomolog zu Hsap P1313) konnten in wiederholten Experimenten keine PCR-Amplifikate dargestellt werden, obwohl RT-PCRs mit dem Forwardprimer P2590 gezeigt haben, daß diese Region im Transkript vorhanden ist. Da auch per Sequenzierung im Bereich von P2589 keine weiteren Splice-ereignisse in den Rind *Ddx3y* Transkripten festgestellt werden konnten, ist davon auszugehen, daß P2589 entweder thermodynamisch nicht optimal für die PCR-Reaktion ist, oder der Primer von Werk aus nicht korrekt synthetisiert wurde.

Mit dem Primerpaar P2591/2567 (P2591 liegt in *MSY2*) konnte ein sehr schwaches PCR-Amplifikat mit einer ungefähren Größe von 250 bp detektiert werden (Abb. 4-30). Per Sequenzierung konnte bestätigt werden, daß es sich hierbei um ein Rind *Ddx3y* Transkript mit *MSY2*-Anteil und einer Länge von 249 bp handelt. Dieses Rind *Ddx3y* Transkript wird innerhalb

von *MSY2* an einer kanonischen 5´GT-Donor-Splicestelle gesplict (siehe Abb. 4-27). Zur Unterscheidung, zu den Human Exon-T 5´Splicesignalen, wird diese Rind Splicestelle **„Exon-Tc“** genannt. Die 3´AG-Spliceakzeptorstelle ist wieder die Rinder-spezifische Exon-1c 3´AG-Splicestelle. Die kurzen Rind *Ddx3y* Transkripte mit Start in der *MSY2* Sequenz entstehen also durch die Rinder-spezifische Splicevariante **Exon-Tc-1c**. Ob dieses *Ddx3y* Transkript allerdings funktionelle Bedeutung für die Genexpression von Rind *Ddx3y* hat, ist auf Grund der sehr geringen Transkriptmengen fraglich. Die Funktionalität dieses Splicingproduktes wird zudem weiter in Frage gestellt durch den Umstand, daß nach Vektor-Klonierung und Sequenzierung des P2591/2567 RT-PCR-Produkt-Pools weitere Splicingprodukte, mit kanonischen und nicht-kanonischen Splicingsignalen, sichtbar wurden, die im Original-Gelbild nicht zu sehen waren. Rind *Ddx3y* Transkripte mit Start im Bereich *MSY2* sind daher scheinbar Ziel alternativer Spliceereignisse, die allerdings zum Teil auch zufällig durch fehlerhaftes, nicht-funktionelles alternatives Splicing generiert werden können (Sorek et al., 2004; Skandalis et al., 2010).

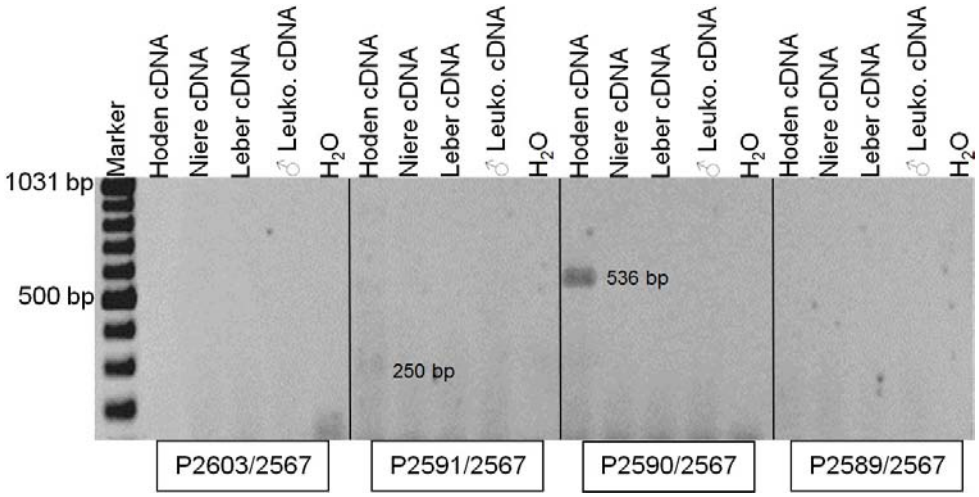


Abb. 4-30: RT-PCR Analyse Rind *Ddx3y*, Exon-T und *MSY2* Region.
Errechnete RT-PCR Amplifikatgrößen für ungesplicte Transkripte: P2589/2567 = 843 bp, P2590/2567 = 934 bp, P2591/2567 = 1031 bp, P2603/2567 = 1081 bp. Das PCR-Produkt mit P2591/2567 ist so schwach, daß es kaum erkennbar ist. DNA-Marker ist eine 100bp DNA-Leiter. (Nach Zimmer J., unveröffentlichte Resultate).

Mit dem Primerpaar P2603/2567 (P2603 liegt vor *MSY2*) konnten keine PCR-Amplifikate mehr erhalten werden. Im Bereich von Rind *MSY2* konnten zwei potentielle Core-Promotermotive lokalisiert werden. Für die Rind Exon-Ta-1c Transkripte konnte direkt vor der P2590 Primer-position ein potentielles Inr-Motiv „TTACGTC“ (ein Mismatch) lokalisiert werden. Stromaufwärts in Position -28 konnte ein potentielles TATA-Motiv „TCATTCAA“ und stromabwärts konnte in Position +28 ein potentielles DPE-Motiv „GAAACT“ identifiziert werden.

Als Startposition für dieses Rind ***Ddx3y* T-TSS-I** Transkript mit Exon-Ta-1c Splicing wird die Position 140.621 im Y-BAC CH240-56A13 vorgeschlagen. Vergleichbar zur Human *DDX3Y* T-TSS-II im *MSY2-2* Repeat nutzt offensichtlich auch das Rind die AT- bzw. CT-reichen *MSY2*

Sequenzen zur Transkriptinitiation. Das TATA-Motiv des vorgeschlagenen Rind T-TSS-I Moduls würde übertragen auf das Human Model in Inr-Block V und das Inr-Motiv zwischen den Inr-Blöcken VI und VII liegen (siehe Abb. 4-3). Das DPE-Motiv „GAAACT“ liegt bereits stromabwärts der *MSY2* Sequenz.

In den Abb. 4-29 u. 4-30 fällt auf, daß mit dem Forwardprimer P2571 wesentlich mehr *Ddx3y* Transkripte amplifiziert wurden, als mit P2590. Ob zwischen den Primerpositionen P2590 und P2571 ein weiterer Core-Promoter aktiv ist, kann per RT-PCR Analyse nicht geklärt werden. Ein vergleichbares Core-Promotermodul zu Human *DDX3Y* T-TSS-I ist auch in der Rindsequenz stromabwärts von *MSY2* kartierbar. So befindet sich positionshomolog zum Human Inr-Motiv die Rind-Sequenz „CTGGACG“. Stromaufwärts ist in Position -32 ein kanonisches TATA-Motiv „AACTTGAA“ und stromabwärts in +18 die Sequenz „CCCGCAGCCTGG“ vorhanden, die, ebenso wie das potentielle Human MTE, fünf Abweichungen zur Konsensussequenz aufweist. Der Unterschied in den Transkriptmengen kann auch daran liegen, daß P2590 direkt am 5´ Transkriptende der *Ddx3y* T-TSS-I lokalisiert ist und somit die PCR-Reaktion thermodynamisch nicht ideal verläuft.
Für die zweite Initiationsstelle vor der P2591 Primerlokalisation konnte direkt stromaufwärts der Rind *MSY2*-Sequenz ein potentielles Inr-Motiv „CCAGTGA“ (+1 Position = 140.742) lokalisiert werden, das mit einem potentiellen TATA-Motiv „ATTACTAG“ an Position +28 ein Core-Promotermodul bilden könnte. Alternativ ist direkt vor P2591 ein potentielles, kanonisches Inr-Motiv „CTAGACC“ (+1 Position = 140.732) für Rind **„*Ddx3y* T-TSS-II“** kartierbar. Die potentiellen Core-Promotermotive für eine Rind *Ddx3y* Exon-T- und *MSY2*-Expression sind in Abb. 4-27 farbig markiert.

Die graphische Zusammenfassung der Rind *Ddx3y* TSS-Analyse ist in Abb. 4-31 zu sehen. Erst nach Abschluß der experimentellen Arbeiten wurde festgestellt, daß die hier verwendeten somatischen Rindergewebe irrtümlich weiblichen Ursprungs waren. Eine Wiederholung der Experimente mit kontrolliert männlichen Geweben zeigte, daß Rind *Ddx3y* Transkripte in der TSS-I Region nicht hodenspezifisch, sondern ebenfalls ubiquitär exprimiert werden (Zimmer J., unveröffentlichte Resultate). Mit den Forwardprimern P2568 und P2569 sind in Hoden, Niere, Leber, Milz und Leukozyten Rind *Ddx3y* Transkripte deutlich darstellbar. Für P2570 konnte eine Expression in Hoden und Leukozyten festgestellt werden (Zimmer J., unveröffentlichte Resultate). Die Expression von Rind Exon-T bleibt, auch nach Wiederholung der Experimente, hoden-spezifisch. Ein aktiver Core-Promoter für die ubiquitäre Rind *Ddx3y* Expression sollte demnach tatsächlich zwischen den P2570 und P2569 Primerlokalisationen zu finden sein. Die ubiquitären Rind *Ddx3y* Transkripte haben also eine 5´UTR-Länge von ~ 100 bp (Pos. von P2569 in BAC-Pos. 139.891).

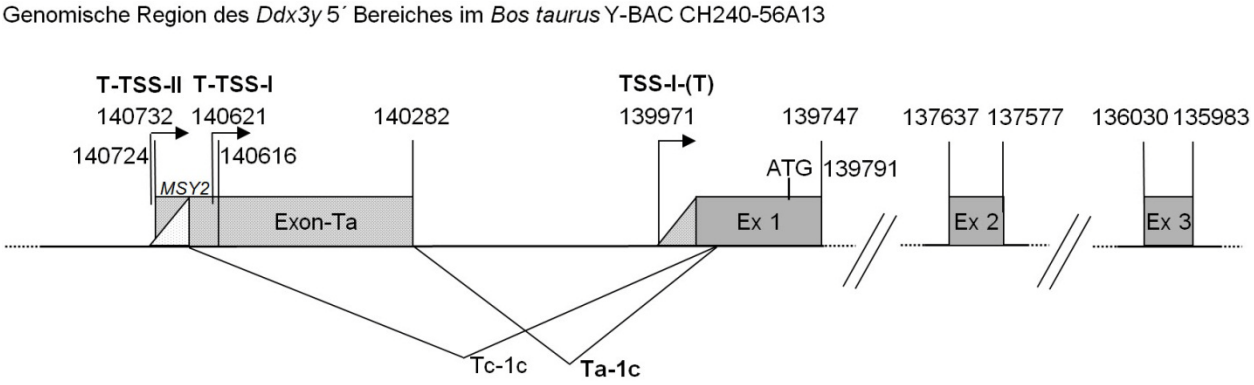


Abb. 4-31: Graphische Zusammenfassung der Rind *Ddx3y* Transkriptionsinitiationsstellen.
Eingezeichnet sind die drei *Ddx3y* TSS Regionen mit jeweiliger Position im Rind Y-BAC CH240-56A13, sowie die Splicevarianten der 5´ verlängerten Transkript mit Start in Rind T-TSS-I und T-TSS-II. Die Rind *MSY2* Sequenz ist schraffiert dargestellt, der Bereich von T-TSS-II bis zur Exon-Tc Splicestelle ist als gepunktetes Dreieck dargestellt. Die Splicevariante Exon-Ta-1c ist in Fettdruck markiert, da es sich hierbei um die Haupt-Splicevariante der Rind *Ddx3y* Transkripte mit Exon-T handelt.

4.5. Evolution der *DDX3X* Keimbahn Core-Promoterstruktur in verschiedenen Säugerspezies

4.5.1 Evolution der *DDX3X* Keimbahn Core-Promoterstruktur in Primaten

Analog zu Kapitel 4.4.1 wird in diesem Abschnitt die Komplexität der *DDX3X* Transkriptions-startstellen in drei Primaten (Schimpanse, *Macaca*, *Callithrix*) im Vergleich zum Human *DDX3X* Gen untersucht.

Als erster Schritt wurden dazu die homologen X-chromosomalen Genomsequenzen per BLAST-Suche aus den Datenbanken extrahiert. Die entsprechenden *DDX3X* Loci konnten für Schimpanse im BAC-Klon „CH251-3A20“[29], für *Macaca* in der X-chromosomalen Sequenz[30] und für *Callithrix* im BAC-Klon „CH259-42H12“[31] lokalisiert werden. Für *Macaca* ist bisher kein BAC im Bereich des *DDX3X* Locus in den Datenbanken zu finden. Dabei handelt es sich jeweils um vollständige *DDX3X* Gene mit identischer Anzahl von 17 Exonen. Die Homologien in den kodierenden Primaten *DDX3X* Sequenzen betragen zwischen 98 und 99%. Die Tabellen mit den Primaten *DDX3X* Exon-Intronstrukturen befinden sich im Anhang (Tab. A-10 bis A-13). Um die kartierten Human *DDX3X* Core-Promoter Module vergleichend zu studieren, wurde, mit dem Alignment-Programm ClustalW, stromaufwärts des Human Translationsstartcodons ATG in Exon1 ein Sequenzalignment der vier *DDX3X* Primaten-Sequenzen erstellt (Abb. 4-32). Das Alignment zeigt, daß die *DDX3X* 5´ Region auf dem X-Chromosom der Primaten sehr stark

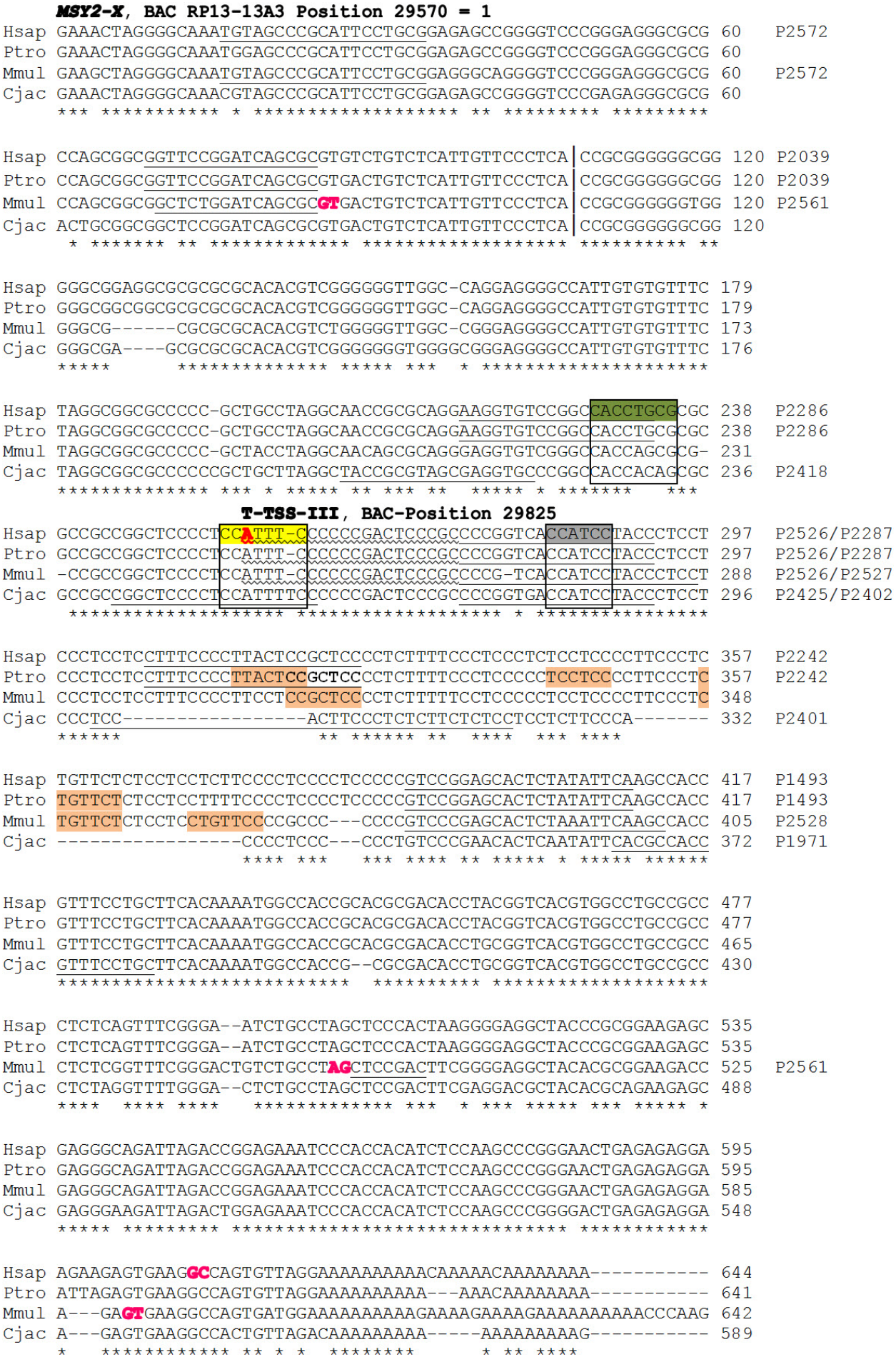
29 Ptro: GenBank Acc.No. AC183643.2
30 Mmul: Ensembl Assembly MMUL 1.0, Datenbankversion 62.10p.
31 Cjac: GenBank Acc.No. AC193908.1; revers-komplementär.

konserviert ist. Zwischen den vier Primatenspezies der Unterordnung Trockennasenaffen (Haplorrhini) ist eine Homologie von über 90% zu der Human-Sequenz, mit Maximum von 98% zwischen Mensch und Schimpanse, festzustellen. Die einzigen Regionen mit deutlichen Sequenzunterschieden sind erstens eine sehr Pyrimidin-reiche Region stromabwärts der *DDX3X* T-TSS-III (fehlt teilweise in *Callithrix*) und zweitens zwei Purin-reiche Regionen stromaufwärts der T/B-TSS-II Position (Abb. 4-32). Diese Sequenzbereiche werden in den Kapiteln 4.6.2 und 4.6.11 detailliert besprochen.

Das vorgeschlagene Human TSS-I Core-Promoter Modul für ubiquitäre *DDX3X* Transkription, bestehend aus TATA-Box, BRE^{d}- und Initiator-Motiv, ist in allen gezeigten Primatenarten vorhanden (Abb. 32). Das Initiator-Motiv „CCAGAGC“ mit der Human TSS-I in BAC-Position 30.639 ist in den vier Spezies zu 100% konserviert. Das TATA-Motiv „TATATTAG“ in Position -32 stromaufwärts ist ebenfalls vollständig konserviert. Das unmittelbar anschließende BRE^{d}-Motiv „ATCCGTG“ ist in *Macaca* identisch wiederzufinden. In *Callithrix* konnte eine neutrale G-Mutation in der ersten Position zu „GTCCGTG“ festgestellt werden. Beim Schimpansen konnte eine negative C-Mutation in der ersten Position zu „CTCCGTG“ gefunden werden, die die Mismatch-Anzahl, in Bezug auf die Konsensussequenz, auf drei erhöht. Die funktionelle Bedeutung dieses *DDX3X* TSS-I Core-Promotermoduls wird somit durch seine Konserviertheit unterstrichen.

Das Core-Promoter Modul der experimentell kartierten Human *DDX3X* T/B-TSS-II in BAC-Position 30.314 ist in den anderen Primatenspezies ebenfalls wiederzufinden (Abb. 4-32). Das Initiator-Motiv „TTAGCCG“ ist im Schimpansen unverändert vorzufinden. Bei *Macaca* und *Callithrix* sind eine bzw. zwei neutrale Mutationen im Inr-Motiv zu finden. Das stromaufwärts gelegene potentielle TATA-Motiv „GAAGAGGC“ in Position -33 ist in allen vier Spezies zu 100% konserviert wiederzufinden. Das kanonische DPE-Motiv „TCTCGC“ in Position +28 ist ebenso vollständig konserviert.

Die dritte, im Humangewebe hodenspezifische, *DDX3X* Initiationsstelle T-TSS-III ist in allen vier Spezies ebenfalls deutlich konserviert (Abb. 4-32). Das kanonische Inr-Motiv „CCATTTC“ liegt in den anderen Altweltaffen (Ptro, Mmul) zu 100% konserviert vor. In *Callithrix* ist die letzte Position, durch Insertion eines Thyminrestes, neutral mutiert. Das potentielle TATA-Motiv „CACCTGCG“ in Position -28 ist in Schimpanse identisch wiederzufinden. Für *Macaca* konnte eine Transversion in Position 5 des TATA-Motives zu „CACCAGCG“ festgestellt werden. In der *Callithrix* TATA-Sequenz „CACCACAG“ sind drei Transversionen in Position 5, 6 und 7 erkennbar. Das in +28 stromabwärts lokalisiert DPE-Motiv „CCATCC“ ist in der Primaten-Evolution von *Callithrix* bis einschließlich Mensch nicht verändert worden.


MSY2-X, BAC RP13-13A3 Position 29570 = 1
T-TSS-III, BAC-Position 29825

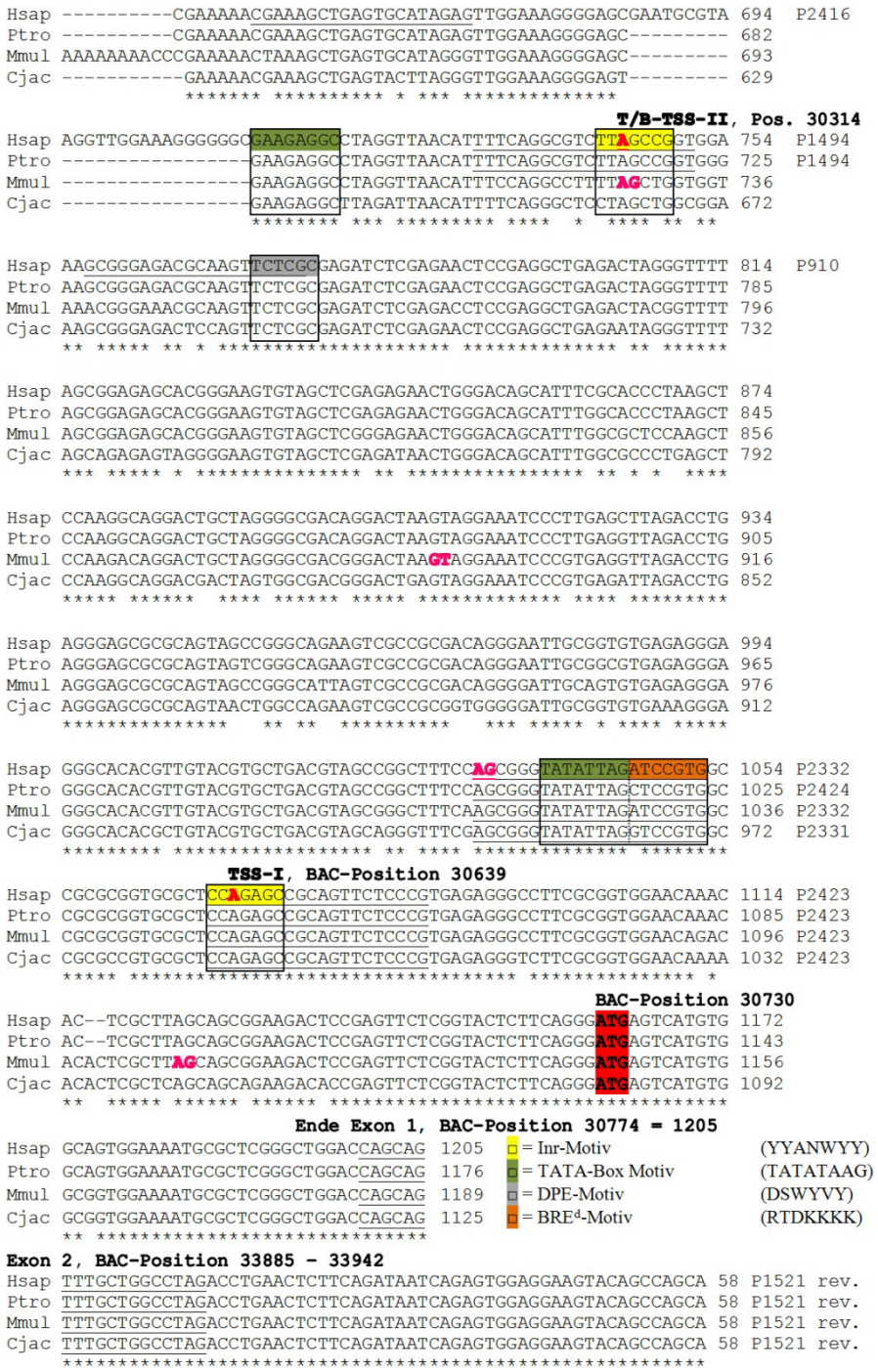


Abb. 4-32: Genomisches *DDX3X* Sequenzalignment zwischen Mensch (Hsap), Schimpanse (Ptro), Rhesusaffe (Mmul) und Weißbüschelaffe (Cjac).

Humansequenz von *MSY2-X* bis Ende Exon1 (Position 29.570 – 30.774; BAC RP13-13A3). Exon2 (BAC-Position 33.885 – 33.942) angegeben zur Darstellung des Reverseprimers P1521. Ptro BAC-Pos. 465 - 1.640 (CH251-3A20), Mmul genomische X-Chromosom Pos. 39.101.003 - 39.102.191 (Mmul Assembly 1.0), Cjac BAC-Pos. 97.841 - 96.717 (CH259-42H12, rev.-komplementär). Kartierte Core-Promotermotive in der Hsap *DDX3X* Sequenz wie in Abb. 4-9. Inr-Motive gelb markiert, TSSs mit roten Buchstaben A hervorgehoben. TATA-Box Motive sind grün, BRE^d-Motive sind orange und DPE-Motive sind grau dargestellt. Die positionshomologen Core-Promotermotive in den vier *DDX3X* 5´ Sequenzen sind gerahmt dargestellt. Potentiell alternative Core-Promotermotive in den Non-Human Primaten sind hellorange markiert. Das ATG Startcodon ist rot hervorgehoben, die Human 5´ und 3´Splicesignale sind in Fettschrift und pink hervorgehoben. In der *Macaca*-Sequenz sind potentielle Splicesignale eines *in-silico* vorhergesagten Transkriptes (GenBank Acc.No. XM_001095294; Abb. 4-38) angegeben. P2526 ist zur Unterscheidung von P2287 gewellt unterstrichen. Die X-spezifische *MSY2-X* Sequenz ist durch einen vertikalen Strich begrenzt. Sterne unterhalb des Alignments geben homologe Nukleotide an.

Die experimentelle Evaluierung der *DDX3X* Transkription in Primatengeweben folgt dem aus Humangewebe bekannten Muster von mehreren RT-PCR Experimenten mit sukzessiv nach 5´ verschobenen Forwardprimern. Für alle durchgeführten RT-PCR Analysen wurde der Exon1/2 überlappende Reverseprimer P1521 gewählt. Die Forwardprimer wurden speziesspezifisch verwendet, wenn die ausgewählten Primersequenzen nicht identisch mit dem Human-Primer waren. Eine gezielte experimentelle Überprüfung der Human *DDX3X* T/B-TSS-II Startstelle in Primaten erfolgte nicht, da Hoden- und Hirngewebe gar nicht, oder nicht in ausreichenden Quantitäten für 5´RACE-Experimente erhältlich waren. Es kann daher in dieser Arbeit nicht geklärt werden, ob diese im Humangewebe hoden- und hirnspezifische T/B-TSS-II auch in anderen Primaten entsprechend genutzt wird.

DDX3X TSS-I RT-PCR Analyse:

Pan troglodytes:

Die *DDX3X* TSS-I RT-PCR Analyse für den Schimpanse ergab mit dem Forwardprimer P2423 eine ubiquitäre *DDX3X* Transkription in allen untersuchten Geweben (Abb. 4-33). Mit dem 36 Nukleotide stromaufwärts gelegenen Forwardprimer P2424 ist nur noch in Hodengewebe eine eindeutige Expression nachweisbar (Abb. 4-33). In den anderen Geweben scheint keine oder nur noch eine Basisexpression stattzufinden. Vergleichbar zu der Human *DDX3X* TSS-I RT-PCR Analyse (Abb. 4-10), ist auch beim Schimpansen eine wiederkehrende Doppelbande in Form eines zweiten PCR-Produktes zu beobachten. Für den Schimpansen sind ebenfalls *DDX3* Pseudogenloci auf Chr. X und 4 bekannt (Chang & Liu, 2010), so daß, wie für Human *DDX3X* auch, eine Primerbindung an offensichtlich transkribierte *DDX3* Pseudogen cDNA, bzw. die Bildung chimärer PCR-Produkte (Smyth et al., 2010; Schütze et al., 2011), als Ursache in Betracht kommen. Das erhaltene RNA-Expressionsmuster ist mit dem Ergebnis der Human *DDX3X* TSS-I Analyse vergleichbar.

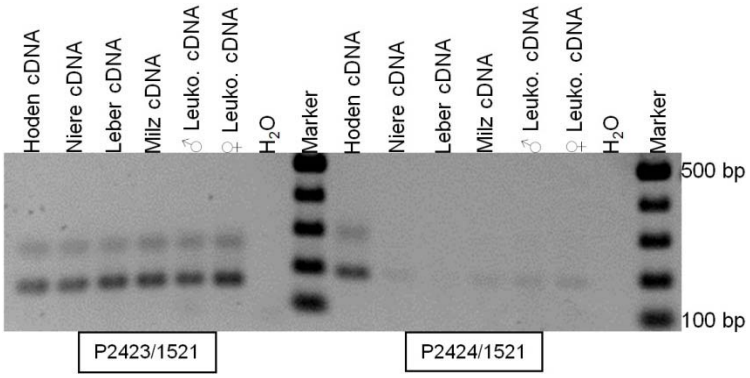


Abb. 4-33: Schimpansen (Ptro) *DDX3X* TSS-I RT-PCR Analyse. PCR-Produkt für P2423/1521 ist 151 bp, für P2424/1521 ist PCR-Produkt 187 bp groß. Untere Bande ist erwartetes *DDX3X* PCR-Produkt, obere ist mögliche Kreuzreaktion mit cDNAs von *DDX3* Pseudogenen oder anderer DEAD-Box Helikasen (siehe Abb. 4-10, Hsap *DDX3X* TSS-I). 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Macaca mulatta:

Die *DDX3X* TSS-I Überprüfung in *Macaca* Geweben bzw. Zellen ergab für den Forwardprimer P2423 ebenfalls eine ubiquitäre *DDX3X* Transkription in den untersuchten Geweben (Abb. 4-34).

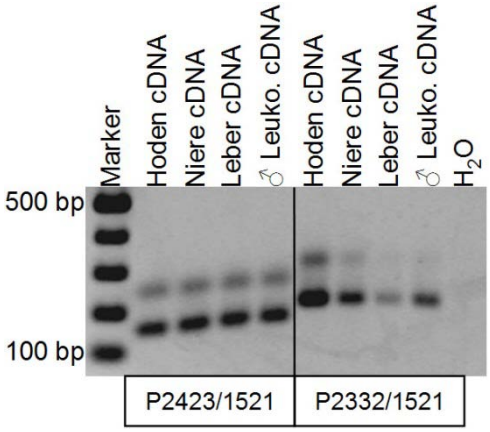


Abb. 4-34: *Macaca* (Mmul) *DDX3X* TSS-I RT-PCR Analyse. PCR-Produkt für P2423/1521 ist 153 bp, für P2332/1521 189 bp groß. Untere Bande ist erwartetes *DDX3X* PCR-Produkt, obere ist mögliche Kreuzreaktion mit cDNAs von *DDX3* Pseudogenen oder anderer DEAD-Box Helikasen (siehe Abb. 4-10, Hsap *DDX3X* TSS-I). Agarosegel ist schräg im elektr. Feld gelaufen. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Im Unterschied zu den Hominiden, Mensch und Schimpanse, ist mit dem Forwardprimer P2332 keine klare Grenze zwischen der ubiquitären TSS-I Region und dem gewebespezifischen 5´ Bereich zu erkennen (Abb. 4-34). In *Macaca* Nierengewebe und in Leukozyten cDNA sind deutliche PCR-Amplifikate mit dem Forward-Primer P2332 zu erkennen. Als mögliche Ursache könnte eine 5´ Expansion des *DDX3X* TSS-I Bereiches für ubiquitäre RNA-Expression in Frage kommen.

Callithrix jacchus:

Die RT-PCR Experimente in Abb. 4-35 zeigen auch für *Callithrix* eine ubiquitäre *DDX3X* Transkription im Bereich von TSS-I. Mit dem Forward-Primer P2423 sind wie bei den Catarrhini PCR-Amplifikate in allen untersuchten Geweben und Zellen zu detektieren. Die stromaufwärts gelegene Region, die mit dem Forward-Primer P2331 untersucht wurde, zeigt ein *DDX3X* RNA-Expressionsmuster wie bei den Hominiden. I.e. 5´ verlängerte *DDX3X* Transkripte sind auch bei *Callithrix*, in funktioneller Menge, nur in Hodengewebe vorzufinden.

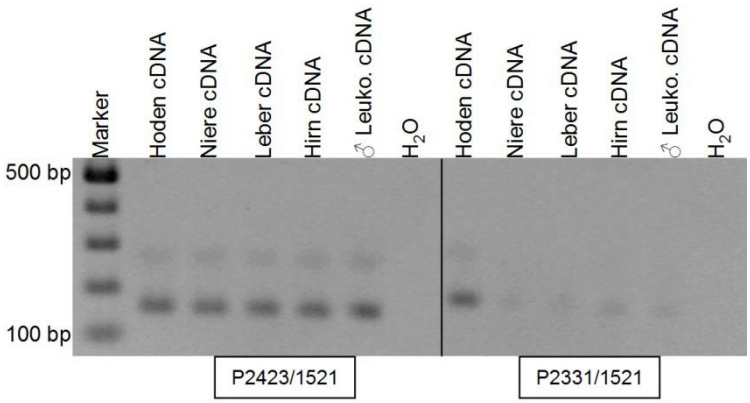


Abb. 4-35: *Callithrix* (Cjac) *DDX3X* TSS-I RT-PCR Analyse. PCR-Produkt für P2423/1521 ist 153 bp, für P2331/1521 ist PCR-Produkt 189 bp groß. Untere Bande ist erwartetes *DDX3X* PCR-Produkt, obere ist mögliche Kreuzreaktion mit cDNAs von *DDX3* Pseudogenen oder anderer DEAD-Box Helikasen (siehe Abb. 4-10, Hsap *DDX3X* TSS-I). 35 PCR-Zyklen. DNA Marker ist eine 100bp DNA-Leiter.

Zusammenfassend kann gesagt werden, daß die in Human bestimmte *DDX3X* TSS-I Region auch in den drei anderen Primatenspezies für eine starke ubiquitäre *DDX3X* RNA-Expression genutzt wird, was auf Grund der fast vollständigen Konserviertheit des *DDX3X* TSS-I Core-Promotermoduls zu erwarten ist. Bei allen drei Spezies weisen die RT-PCR Experimente in der *DDX3X* TSS-I Region eindeutig darauf hin, daß 5´ verlängerte *DDX3X* Transkripte existieren.

DDX3X TSS-III RT-PCR Analyse:

Die experimentellen Überprüfungen der verlängerten *DDX3X* Transkripte mit potentiellem Start in T/B-TSS-II oder T-TSS-III Position ergaben für die drei untersuchten Primaten *DDX3X* Transkripte, die stromaufwärts der T/B-TSS-II Position initiieren. Da in diesem Ansatz die absolute 5´ Grenze der *DDX3X* Transkripte analysiert werden sollte, wurde die potentielle Initiation in der *DDX3X* T/B-TSS-II Position in Primaten nicht weiter verfolgt.

Pan troglodytes:

Die Negativkontrolle mit dem Forwardprimer P2286 ergab für den Schimpansen kein PCR-Produkt, so daß auch hier eine Transkriptinitiation stromaufwärts zur *DDX3X* T-TSS-III Position ausgeschlossen werden kann (Abb. 4-36). Der Forwardprimer P2526, der in der *DDX3X* T-TSS-III Position beginnt, ergab verglichen mit dem Human RT-PCR Experiment, nur ein schwaches PCR-Amplifikat der richtigen Größe (Abb. 4-36). Die RT-PCR mit dem direkt stromabwärts lokalisierten Forward-Primer P2278 ergab ebenfalls ein sehr schwaches PCR-Amplifikat. Mit dem Forward-Primer P2242 konnte kein Produkt detektiert werden. Weiter stromabwärts sind die Forward-Primer P1493 und P1494 lokalisiert mit denen starke hodenspezifische PCR-Amplifikate zu erhalten waren (Abb. 4-36).

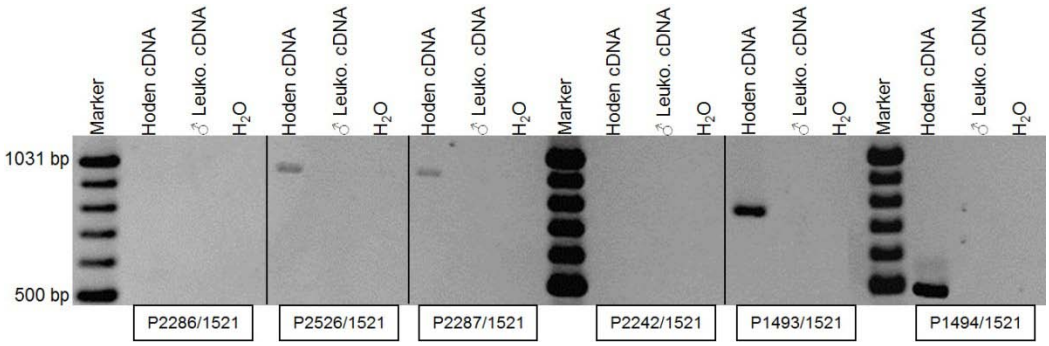


Abb. 4-36: Schimpanse (Ptro) *DDX3X* T-TSS-III RT-PCR Analyse.
Errechnete PCR-Produktgrößen: P2286/1521 = 974 bp, P2526/1521 = 934 bp, P2287/1521 = 915 bp, P2242/1521 = 884 bp, P1493/1521 = 800 bp, Primer P1494/1521 = 487 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Dieser Befund stellt einen Unterschied zum Human *DDX3X* T-TSS-III Ergebnis dar, bei dem auch mit P2526 eine eindeutige RNA-Expression festgestellt werden konnte. Entweder handelt es sich hierbei um ein technisches Problem, daß die cDNA aus Schimpansen Hodengewebe

nicht die Qualität für lange Transkripte besitzt. Es besteht auch die Möglichkeit, daß der Schimpanse die bekannte *DDX3X* T-TSS-III Position nicht gleichermaßen oder gar nicht nutzt und an einer potentiell stromabwärts gelegenen TSS initiiert. Es fällt aber auf, daß die Primer mit schwacher PCR-Reaktion innerhalb der Pyrimidin-reichen Region zu Beginn des *DDX3X* Transkriptes lokalisiert sind, so daß die Primerbindung, bei ohnehin geringer Transkriptmenge, eventuell nicht effizient ablaufen kann. Stromaufwärts zu P1493 konnten drei Inr-Motive „CTGTTCT“, „CCGCTCC“ und „TTACTCC“ lokalisiert werden. Das dritte dieser potentiellen Inr-Motive liegt innerhalb der genomischen Sequenz des Forwardprimers P2242. In Position +28 zu diesem Inr-Motiv ist zudem ein fast-konsensus DPE-Motiv „TCCTCC“ mit einem Mismatch zu identifizieren. Diese potentiellen Schimpansen *DDX3X* Core-Promoter-Motive sind in Abb. 4-32 markiert.

Macaca mulatta:

Die Analyse in *Macaca*-Geweben ergab ein ähnliches Bild wie für den Schimpansen (Abb. 4-37). Mit den Forwardprimern P2526, P2527 und P2242 konnten nur wenige *DDX3X* Transkripte amplifiziert werden (Abb. 4-37). Die Überprüfung in *Macaca* Leber- u. Nierengewebe, die auf Grund der positiven Ergebnisse für somatische Gewebe mit P2332 in Abb. 4-34 nötig wurde, ergab keine PCR-Amplifikate (hier nicht gezeigt). D.h. also, daß auch für *Macaca* gilt, daß Transkripte mit Start in der *DDX3X* T-TSS-III Region ausschließlich in Hodengewebe transkribiert werden.

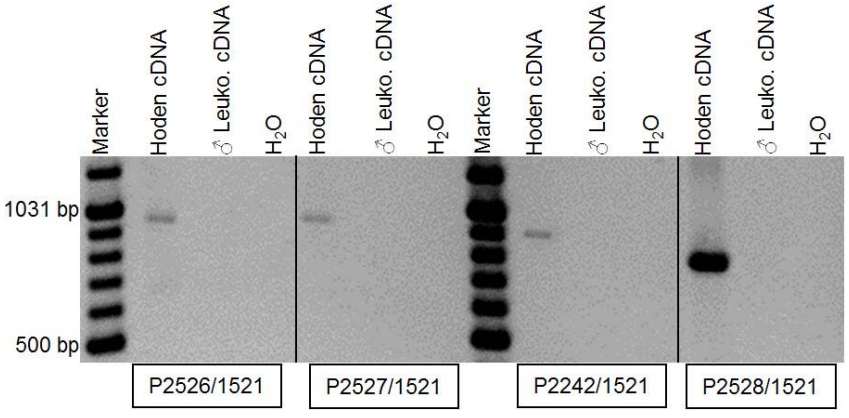


Abb. 4-37: *Macaca* (Mmul) *DDX3X* T-TSS-III RT-PCR Analyse.
Errechnete PCR-Produktgrößen: P2526/1521 = 955 bp, P2527/1521 = 936 bp, P2242/1521 = 906 bp, P2528/1521 = 825 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Mit dem stromabwärts gelegenen Primer P2528 (positionshomolog zu P1493) sind, wie beim Schimpansen, mengenmäßig signifikante PCR-Produkte erst nach der Pyrimidin-reichen Region zu sehen (Abb. 4-37). Alternative Initiationsstellen zur *DDX3X* T-TSS-III Position sind bei *Macaca*, wie zuvor beim Schimpansen, ebenfalls stromaufwärts von P2528 möglich. Zwei der drei potentiellen Inr-Motive „CTGTTCC“, „CTGTTCT“ und „CCGCTCC“ sind identisch mit denjenigen, die in der homologen Schimpansen-Region lokalisiert werden konnten. Das Inr-Motiv mit potentiell assoziiertem DPE-Motiv ist jedoch bei *Macaca*, auf Grund einer Mutation in

der +1 Position des Inr-Motives, nicht mehr funktionell. Die potentiellen *Macaca DDX3X* Core-Promoter-Motive sind in Abb. 4-32 markiert.

Callithrix jacchus:

Die *Callithrix jacchus DDX3X* RT-PCR Analyse ergab auch für diesen Primaten keine *DDX3X* Transkriptionsinitiation stromaufwärts der T-TSS-III Position (P2418, P2425; Abb. 4-38). Mit den Forward-Primern P2526, P2402, P2401 und P1971 konnten jeweils positive PCR-Reaktion mit PCR-Amplifikaten der richtigen Größen beobachtet werden (Abb. 4-38). Wie auch schon zuvor für Schimpanse und *Macaca* registriert, ergaben die PCR-Reaktionen mit den Forwardprimern, die innerhalb der Pyrimidin-reichen Region lokalisiert sind, schwächere Amplifikate als mit den stromabwärts gelegenen Primern (hier P1971). Eine weitere Überprüfung in *Callithrix* Leber- und Nierengewebe bestätigte auch für diese Primatenspezies die Hodenspezifität der 5´ verlängerten *DDX3X* Transkripte.

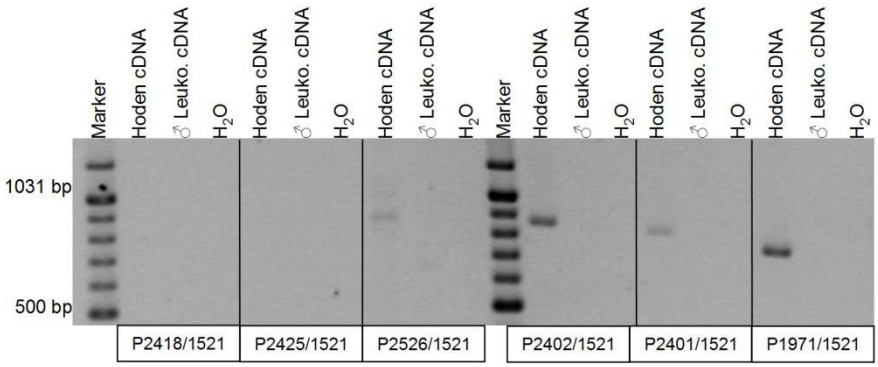


Abb. 4-38: *Callithrix* (Cjac) *DDX3X* T-TSS-III RT-PCR Analyse.
Errechnete PCR-Produktgrößen: P2418/1521 = 936 bp, P2425/1521 = 897 bp, P2526/1521 = 885 bp, P2402/1521 = 865 bp, P2401/1521 = 839 bp, P1971/1521 = 775 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Mit der vorliegenden Analyse konnte eindeutig gezeigt werden, daß in allen vier untersuchten Primatenspezies 5´ verlängerte *DDX3X* Transkripte mit Hodenspezifität transkribiert werden. Die Ergebnisse weisen darauf hin, daß für alle untersuchten Primaten der absolute 5´ Transkriptionsstart in der *DDX3X* T-TSS-III Position lokalisiert ist und das entsprechende Core-Promotermodul funktionell konserviert blieb. Die RT-PCR-Analysen in *Pan* und *Macaca* zeigen aber auch, daß potentielle Core-Promoterelemente stromabwärts der *DDX3X* T-TSS-III Position in der Pyrimidin-reichen Sequenz in diesen Spezies präferentiell genutzt werden können. Die *DDX3X* TSS-I Region ist funktionell hochkonserviert innerhalb der verschiedenen Primaten-spezies.

Es konnte festgestellt werden, daß das hodenspezifische 5´UTR Splicing des Human *DDX3X* Transkriptes nicht in anderen Primatenspezies wiederzufinden ist. Das 5´UTR Splicing stellt demnach eine humanspezifische Eigenschaft dar.

4.5.2. Experimentelle Überprüfung eines computergenerierten *M. mulatta DDX3X* Transkriptes

Bei dem *Macaca mulatta DDX3X*-Transkript „XM_001095294“ aus der NCBI-Datensammlung handelt es sich um ein computer-vorhergesagtes Transkript, das durch RT-PCR Analysen experimentell überprüft werden soll. Das Transkript wurde durch die NCBI-Programme „Gnomon“ und „Splign“, beides Vorhersageprogramme für eukaryotische Gene, annotiert (Kapustin et al., 2008).

Im Unterschied zum Human *DDX3X* Transkript sollte dieses *Macaca DDX3X*-Transkript eine mehrfach gesplicte 5´UTR haben und sollte 5´ deutlich länger sein (Abb. 4-39). Vor der Human *DDX3X* T-TSS-III Position sind zwei weitere „5´UTR-Exone“ vorhergesagt. Die bekannte *DDX3X* 5´UTR-Verlängerung des hodenspezifischen Human-Transkriptes sollte bei diesem *Macaca* Transkript in drei Exone unterteilt sein.

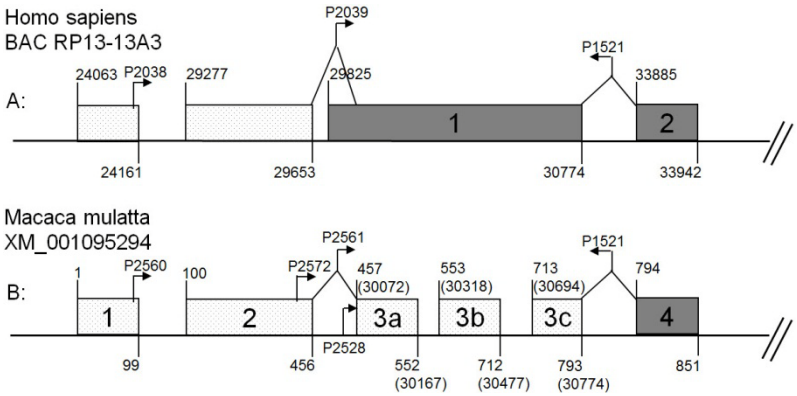


Abb. 4-39: Schematische Darstellung des *in-silico* vorhergesagten *Macaca mulatta DDX3X* Transkriptes „XM_001095294“.

A: Human *DDX3X* 5´UTR Darstellung incl. der zwei zusätzlichen *Macaca DDX3X* 5´UTR-Exone. **B:** Darstellung des vorhergesagten *Macaca DDX3X* Transkriptes. Die Positionen der Splicesignale sind im Sequenzalignment Abb. 4-32 markiert. Die Positionsangaben geben die Länge der 5´UTR an. Die Positionsangaben in Klammern sind in Bezug zum Human BAC RP13-13A3 angegeben.

Die Überprüfung per RT-PCR erfolgt mit vier Primerkombinationen, die alle den Primer P1521 aus Exon2 als gemeinsamen Antisenseprimer nutzen. Die Forwardprimer P2560 und P2572 liegen jeweils in einem der vorhergesagten 5´UTR-Exone. Der Forwardprimer P2561 ist überlappend für das 5´UTR-Exon 2 und 3a entwickelt worden (Abb. 4-39). Der Forwardprimer P2528, der entsprechend der *in-silico* Transkriptvorhersage im Intron zwischen 5´UTR Exon 2 und 3a liegen würde, dient als Kontrolle und zeigt ein 825 bp langes PCR-Amplifikat in cDNA aus Hodengewebe (Abb. 4-40). In Abb. 4-40 ist zu erkennen, daß für keine der PCR-Reaktionen, die mit den spezifischen Oligonukleotiden für das vorhergesagte *Macaca DDX3X* Transkript durchgeführt wurden, ein PCR-Produkt zu detektieren war.

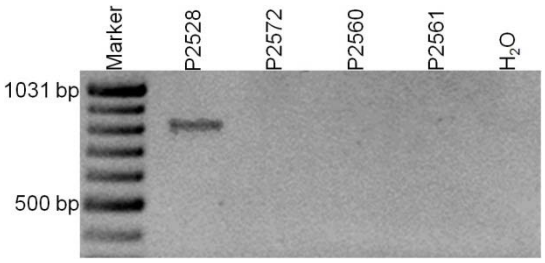


Abb. 4-40: RT-PCR Analyse des vorhergesagten *Macaca mulatta DDX3X* Transkriptes XM_001095294. Die hier gezeigten PCR-Reaktionen sind mit cDNAs aus *Macaca* Hodengewebe durchgeführt. Errechnete PCR-Produktgrößen: P2528/1521 = 825 bp für durchgehendes Transkript, P2572/1521 = 420 bp für gesplictes Transkript, P2560/1521 = 764 bp für gesplictes Transkript, P2561/1521 = 366 bp für gesplictes Transkript. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

In Abb. 4-37 ist ebenfalls zu sehen, daß P2528/1521 eine positive RT-PCR in Hodengewebe zeigt. Die RT-PCR-Analyse wurde ebenfalls mit *Macaca* Nierengewebe durchgeführt (hier nicht gezeigt), was ebenso zu einem negativen Ergebnis führte. Auf Grund dieser RT-PCR Analyse ist dieses *in-silico* vorhergesagte *Macaca mulatta DDX3X* Transkript nicht real, da keine experimentelle Evidenz für dessen Existenz vorhanden ist.

4.5.3. Evolution der *DDX3X*/*Ddx3x* Keimbahn Core-Promoterstruktur in Rind und Maus

Nachdem in Kapitel 4.5.1 aufgezeigt werden konnte, daß die 5´UTR Verlängerung des *DDX3X* Transkriptes in der Keimbahn nicht humanspezifisch, sondern bereits in den untersuchten Primatenspezies vorhanden ist, soll nun unter Einbeziehung weiterer Säugerspezies, Maus und Rind, überprüft werden, ob die hodenspezifische 5´UTR Verlängerung nur innerhalb der Primaten konserviert ist, oder ob es sich um eine stärker konservierte Eigenschaft von *DDX3X* Transkripten in der Keimbahn von Säugern handelt.

Ausgangspunkt der vergleichenden Analyse war die Ermittlung der korrespondierenden genomischen *Ddx3x* Loci in Maus und Rind. Mittels BLAST-Suche, gegen die Human *DDX3X* Sequenz aus BAC RP13-13A3 (Position 29.570 – 30.774), konnten in den NCBI-Datenbanken die entsprechenden genomischen Maus und Rind *Ddx3x* Sequenzen identifiziert werden. Für die Maus wurde der *Ddx3x* Locus im BAC-Klon „RP23-158L19“[32] lokalisiert. Für das Rind konnte die entsprechende *Ddx3x* Sequenz in dem genomischen Contig „004.81.scaffold1“[33] lokalisiert werden. Der dazugehörige X-chromosomale BAC-Klon „CH240-81E14“[34] des Rind *Ddx3x* Locus wurde erst nach Abschluß dieser Arbeit in der NCBI-Datenbank verfügbar. Analog zu der *DDX3X* Analyse in Primaten, wurde die potentielle 5´UTR Region von Maus und Rind *Ddx3x* per ClustalW-Sequenzalignment mit der Human *DDX3X* Sequenz verglichen (Abb. 4-41).

---

32 Mmus: GenBank Acc.No. AL833805.11; Position 89.571 - 90.583.

33 Btau: GenBank Acc.No. NW_001501904.1; Position 82.419 – 83.588, rev. komplementär; Assembly Btau 4.0, Datenbankversion 57.4f.

34 Btau: GenBank Acc.No. AC163194.3.

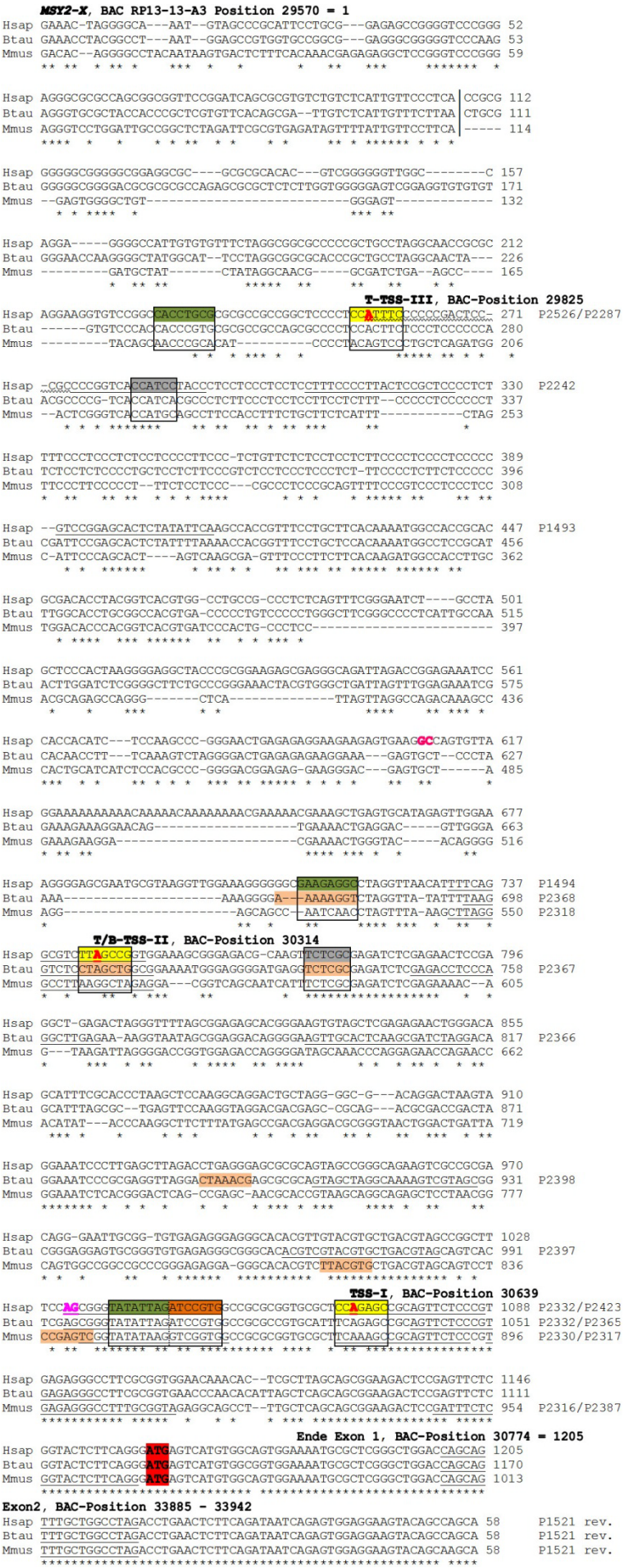


Abb. 4-41: Genomisches *DDX3X/ Ddx3x* Sequenzalignment zwischen Mensch (Hsap), Rind (Btau) und Maus (Mmus).

Genomische Human *DDX3X* Sequenz von *MSY2-X* bis Ende Exon1 (Position 29.570 – 30.774; BAC RP13-13A3). Exon2 (BAC-Position 33.885 – 33.942) angegeben zur Darstellung des Reverseprimers P1521. Btau genomisches Contig „004.81“, Position 82.419 – 83.588 rev. komplementär (GenBank Acc.No. NW_001501904.1; Assembly Btau 4.0, Datenbankversion 57.4f), Mmus BAC-Pos. 89.571 - 90.583 (RP23-158L19). Kartierte Core-Promoter Motive in der Hsap *DDX3X* Sequenz wie in Abb. 4-9. Inr-Motive gelb markiert, TSSs mit roten Buchstaben A hervorgehoben. TATA-Box Motive sind grün, BRE^{d}-Motive sind orange und DPE-Motive sind hellgrau dargestellt. Die potentiell positionshomologen Core-Promotermotive in den drei *DDX3X/Ddx3x* 5´ Sequenzen sind gerahmt dargestellt. Potentiell alternative Maus und Rind Core-Promotermotive sind in hellorange markiert. Das ATG Startcodon ist rot hervorgehoben, die Human 5´ und 3´Splicesignale in der 5´UTR sind in Fettschrift und pink hervorgehoben. Primersequenzen sind unterstrichen und P2526 ist zur Unterscheidung von P2287 gewellt unterstrichen. Die Primerbezeichnungen beziehen sich auf die interne Labornomenklatur. Die X-spezifische *MSY2-X* Sequenz, homolog zu *DDX3Y MSY2* Repeat 1, ist am 3´ Ende durch einen vertikalen Strich markiert. Sterne unterhalb des Alignments geben homologe Nukleotide an.

Anhand des Alignments wurden dann speziesspezifische Oligonukleotide entwickelt, um die mRNA-Expression von *Ddx3x* per RT-PCR in Maus- und Rindergeweben nachweisen zu können. In Abbildung 4-41 ist deutlich sichtbar, daß die Sequenzhomologie zwischen Mensch, Rind und Maus in der X-chromosomalen Linie wesentlich höher ist, als dies für die positionshomologen Y-chromosomalen *DDX3Y/Ddx3y* Bereiche der Fall ist (Abb. 4-27). Trotz der geringeren evolutionären Distanz zwischen Mensch und Maus (Abb. 1-5) ist für die betrachtete *DDX3X/Ddx3x* Promoterregion eine geringere Sequenzhomologie von ~ 60%, im Vergleich zu ~ 75% zwischen Mensch und Rind, festzustellen (Abb. 4-41). Im proximalen Promoterbereich weisen alle drei Spezies ein hohes Maß an Sequenzhomologie auf. Im distalen Sequenzbereich des Alignments fällt vor allem in der Maussequenz auf, daß mehrere indel-Ereignisse im Lauf der Evolution erfolgt sein müssen und so zu einer reduzierten Sequenzhomologie führten. Die Homologien in den kodierenden Maus und Rind *Ddx3x* Sequenzen betragen 93% für die Maus und 95% für das Rind. Die Tabellen mit den Rind und Maus *Ddx3x* Exon-Intronstrukturen befinden sich im Anhang (Tab. A-14, A-15).

Maus *Ddx3x* Core-Promoter & RT-PCR Analyse:

Für Maus *Ddx3x* konnten TSS- und Gewebeinformationen in der DBTSS ermittelt werden, die in Tabelle 4-6 zusammengefaßt sind. Von acht kartierten Einträgen aus der DBTSS enthalten nur zwei das Y-R-Motiv. Keiner der Einträge entspricht einem perfekten Inr-Motiv. Da nur die TSS-Positionen mit Y-R-Dinukleotid berücksichtigt werden, umfaßt dieser Maus *Ddx3x* TSS-Bereich 6 Nukleotide. Auffällig ist dennoch, daß die zwei Positionen (BAC-Positionen 90.445 und 90.446) ohne Y-R-Dinukleotid die meisten DBTSS-Nennungen auf sich vereinen.

Drei weitere potentielle TSS-Positionen konnten aus den NCBI-und Ensembl-Datenbanken extrahiert werde. Die Maus *Ddx3x* Transkripte aus NCBI (GenBank Acc.No. NM_010028.3) und aus Ensembl (ENSMUST00000153851) beginnen in BAC-Position 90.448. Wenn auch nicht sicher sein kann, daß diese Transkripte 5´ vollständig sind, entspricht die Maus BAC-Position 90.448 der Human *DDX3X* TSS-I Position. Die zwei *Ddx3x* Transkripte aus der Ensembl-Datenbank (ENSMUST00000000804 u. ENSMUST00000153611) beginnen in Maus BAC-Position 90.396 bzw. 90.458. Der Transkriptbeginn in BAC-Pos. 90.396 gibt einen ersten Hinweis auf eine möglicherweise gewebespezifische 5´ Transkriptverlängerung, da diese Position ~ 50 Nukleotide stromaufwärts der in DBTSS erfaßten ubiquitären TSSs lokalisiert ist.

Das Human *DDX3X* TSS-I Core-Promoter Modul ist in der Maussequenz positionshomolog wiederzufinden (Abb. 4-41). Im Maus TSS-I Inr-Motiv „TCAAAGC“ sind im Vergleich zum Human Inr-Motiv „CCAGAGC“ zwei neutrale Mutationen zu finden. Das stromaufwärts gelegene Maus TATA-Motiv „TATATAAG“ entspricht perfekt der Konsensussequenz und im BRE^{d}-Motiv

„GTCGGTG“ ist eine neutrale und eine positive Mutation gegenüber der Humansequenz „ATCCGTG“ festzustellen. Das optimale TATA-Motiv in der Maus *Ddx3x* Sequenz scheint die Positionierung des PIC zu regulieren, da das TATA-Motiv in Position -29 bzw. -30 bezüglich der TSS-Positionen ohne Y-R-Dinukleotid, aber mit den meisten DBTSS-Nennungen lokalisiert ist (BAC-Pos. 90.445 und 90.446). Eine solche Eingrenzung der TSS-Positionen läßt auf einen fokussierten Core-Promotertypus mit starker TATA-Box schließen.

Tab. 4-6: Positionen potentieller Initiator-Motive im Bereich des ubiquitären Maus *Ddx3x* Transkriptstartes in Exon1 (TSS-I Region).

Mismatches der ermittelten Sequenzen zur Inr-Konsensussequenz sind rot markiert. Das hochkonservierte Y-R Dinukleotid ist in Spalte 2 unterstrichen. Initiator-Motive mit 100% Übereinstimmung zur Konsensussequenz sind im Fettdruck und grün hervorgehoben. Weitere potentielle TSS mit assoziiertem Y-R Dinukleotid sind hellgrau dargestellt. TSSs die nicht der Inr-Seq. entsprechen bzw. kein Y-R Dinukleotid enthalten, sind farblich nicht unterlegt. Die Positionsangaben entsprechen der +1 Position (= TSS) des Transkriptes, die wiederum dem dritten Nukleotid in der Inr-Sequenz entspricht. In Spalte 4 wird angegeben aus welchen Geweben/Zellen die Transkripte isoliert wurden. Die Zahlenangaben in Spalte 5 geben die Anzahl der Nennungen in der DBTSS an. Positionsangaben beziehen sich auf den murinen X-BAC RP23-158L19. Maus *Ddx3x* ATG ist an BAC-Position 90.539.

Position der 5´Enden der DBTSS Klone	Sequenz des pot. Inr Elementes: YYANWYY	Y-R Regel für -1/+1 Position im Inr	Gewebe und Anzahl cDNA-Klone pro Gewebe	Gesamtzahl cDNA-Klone in DBTSS
90.443	GCGCTTC	+	Blastozysten, Total-Embryo (2)	3
90.445	GCTTCAA	-	Embryo, 8-Zell-Stadium, Hirn, Total-Embryo (9)	12
90.446	CTTCAAA	-	Embryo-Kopf, 8-Zell-Stadium, B16-F10Y Melanoma-Zellinie, Milchdrüse (3), Blastozysten, Lunge (2), Total-Embryo (7), Stromazellen	17
90.447	TTCAAAG	-	Embryo-Ovar (2), Milchdrüse (3), Hirn, Synovialfibroblasten	7
90.448	TCAAAGC	+	Total-Embryo, Sugano-Maus Niere (3)	4
90.449	CAAAGCC	-	Innenohr	1
90.464	CCCGTGA	-	Sugano-Maus Niere, Sugano-Maus Leber	2
90.475	GCCTTTG	-	2-Zell-Stadium	1

Die kartierte Human *DDX3X* T/B-TSS-II Position (BAC-Position 30.314) mit dem Inr-Motiv „TTAGCCG“ konnte nicht positionshomolog in der genomischen Maussequenz identifiziert werden. Durch mehrere Sequenzveränderungen entspricht die Sequenz „AAGGCTA“ nicht mehr dem Inr- und Y-R-Motiv. Ein putatives Inr-Motiv der Sequenz „TTAAGGC“ ist allerdings unmittelbar stromaufwärts kartierbar. Das stromabwärts in Position +28 gelegene DPE-Motiv „TCTCGC“ ist zu 100% konserviert. Stromaufwärts ist positionshomolog in Position -30 ebenfalls ein potentielles TATA-Motiv „CAATCAAC“ zu finden.

Für die Human *DDX3X* T-TSS-III Core-Promotermotive konnten in der genomischen Maus-sequenz ebenfalls entsprechende Elemente identifiziert werden. Das positionshomologe, potentielle Maus Inr-Motiv „ACAGTCC“ und das konservierte DPE-Motiv „CCATGC“ sind in der

Maus *Ddx3x* Sequenz kartierbar. Das potentielle TATA-Motiv ist auf Grund von Sequenz-mutationen und indel-Ereignissen nicht konserviert.

Die experimentelle Überprüfung der Maus *Ddx3x* TSS-Region erfolgte mit gezielten RT-PCR Experimenten. Die *Ddx3x* RT-PCR Analyse in Maus-Hodengewebe und in männlichen Leukozyten ist in Abb. 4-43 zu sehen. Die Forwardprimer P2316 und P2317 markieren die Maus TSS-I Region und bestätigen die DBTSS und Core-Promoter Daten.

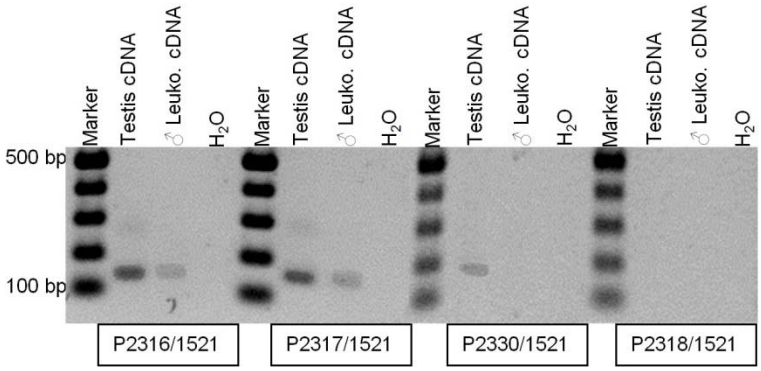


Abb. 4-43: RT-PCR Analyse Maus *Ddx3x*.
Errechnete PCR-Amplifikatgrößen für P2316/1521 = 131 bp, P2317/1521 = 151 bp, P2330/1521 = 185 bp, P2318/1521 = 482 bp. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Mit dem stromaufwärts gelegenen Forwardprimer P2330 konnte ein schwaches PCR-Produkt nur in Maus-Hodengewebe erhalten werden (Abb. 4-43). Der Beginn des Ensembl-Transkriptes (ENSMUST00000000804) mit Start in BAC-Position 90.396 liegt direkt stromaufwärts der P2330 Primerposition. Zusammen mit dem RT-PCR Ergebnis für Primer P2330 ergibt dies einen Hinweis für eine hodenspezifische Maus *Ddx3x* 5´UTR-Transkriptverlängerung mit einer 5´UTR-Länge von mind. 127 bp (bis einschließlich P2330 Position berechnet). Zwei mögliche Inr-Motive für ein solches *Ddx3x* Transkript sind in Abb. 4-41 markiert. Mit dem Forwardprimer P2318, der im Bereich einer potentiellen TSS-II lokalisiert ist, konnte kein RT-PCR Produkt mehr festgestellt werden (Abb. 4-43).

Die *Ddx3x* RNA-Expression in Maus-Hodengewebe, aber vor allem in Leukozyten, war in allen Experimenten wesentlich geringer als vergleichbar für Maus *Ddx3y* gezeigt (Abb. 4-28). Die geringe *Ddx3x* RNA-Expression ist daher kein Problem der cDNA Qualität, da die identischen cDNA Proben für die *Ddx3x* wie *Ddx3y* RT-PCR-Analyse verwendet wurden.

Rind *Ddx3x* Core-Promoter & RT-PCR Analyse:

Für die Rind *Ddx3x* Analyse standen keine DBTSS-Daten zur Verfügung, so daß die RT-PCR Analyse anhand der bekannten Human *DDX3X* RT-PCR Ergebnisse erfolgte.

Aus der NCBI-Datenbank war ein Rind *Ddx3x* Transkript „NM_001192962.1“ extrahierbar. Dieses Rindertranskript beginnt in Position 82.556 des genomischen Contigs „004.81“ (rev.-komplementär), welche der Human *DDX3X* TSS-I Position in BAC-Position 30.639 entspricht.

Das dominante Inr-Motiv der Human *DDX3X* TSS-I Region ist positionshomolog in der genomischen Rindsequenz wiederzufinden (Abb. 4-41). Das Inr-Motiv „TCAGAGC“ weist eine funktionell neutrale Mutation im Vergleich zur Human Inr-Sequenz auf. Stromaufwärts sind das TATA- und BRE^{d}-Motiv zu 100% konserviert. Das *DDX3X/Ddx3x* TSS-I Core-Promotermodul, aus TATA-Box, BRE^{d}- und Inr-Motiv, scheint also bereits seit über 100 Mio. Jahren von Rind bis Mensch konserviert zu sein und unterstreicht die funktionelle Bedeutung dieses Moduls.
Eine *Ddx3x* TSS-II Region ist potentiell auch für das Rind möglich. Das Inr-Motiv „CTAGCTG“, mit zwei funktionell neutralen Mutationen im Vergleich zum Human Inr-Motiv, ist homolog zum Human Inr-Motiv in Position T/B-TSS-II (Hsap BAC-Position 30.314) zu finden. Das vollständig konservierte DPE-Motiv „TCTCGC“ ist um eine Position verschoben und ist nun in der nicht optimalen Position +29 lokalisiert. Das stromaufwärts gelegene positionshomologe TATA-Motiv in Position -32 hätte bei Rind die Sequenz „AAAAAGGT“.
Ein Rind *Ddx3x* TSS-III Core-Promoter ist ebenfalls positionshomolog kartierbar. Das potentielle Rind TSS-III Inr-Motiv „CCACTTC“ entspricht der Konsensussequenz und das stromabwärts lokalisierte DPE-Motiv „CCATCA“ weist ein Mismatch mehr auf, als das DPE-Motiv im Human *DDX3X* T-TSS-III Core-Promoter. Das DPE ist um eine Position an +29 verschoben. Stromaufwärts in Position -28 ist das potentielle, positionshomologe TATA-Motiv „CACCCGTG“ mit zwei Mismatches in den funktionell kritischen Positionen 4 und 5 zu finden.

Die experimentelle Evaluierung der Rind *Ddx3x* mRNA-Expression, mit unterschiedlich langen 5´UTR-Sequenzen, erfolgte wiederum durch gezielte RT-PCR Experimente (Abb. 4-44). Mit dem Forwardprimer P2365 konnte in den drei untersuchten Geweben (Hoden, Niere, Leber) und in Leukozyten eine *Ddx3x* RNA-Expression dargestellt werden (Abb. 4-44). Wie bereits für die Rind *Ddx3y* RT-PCR-Analyse beschrieben, handelt es sich bei den somatischen Rindergeweben fälschlicherweise um weibliches und nicht um männliches Gewebe. P2365 ist direkt stromabwärts der TSS-I Position lokalisiert. Für den stromaufwärts der TSS-I Position gelegenen Forwardprimer P2332 konnte bereits keine ubiquitäre, starke *Ddx3x* Transkription mehr dargestellt werden. Nur in Hodengewebe und in ♀Leukozyten waren größere Transkriptmengen darstellbar. Es ist also davon auszugehen, daß die von Human und Maus bekannte *DDX3X/Ddx3x* TSS-I Position auch in Rind funktionell konserviert vorliegt und für eine ubiquitäre RNA-Expression verantwortlich ist.
Weiter stromaufwärts zeigen die Forwardprimer P2397 und P2398 nur noch PCR-Produkte in Hodengewebe und ♀Leukozyten. Mit den Forwardprimern P2366 und P2367 waren nur noch im

Rind-Hodengewebe PCR-Amplifikate zu detektieren. Mit dem Forwardprimer P2368 konnte in keinem der untersuchten Geweben und Zellen eine *Ddx3x* Transkript-Amplifikation festgestellt werden (Abb. 4-44).

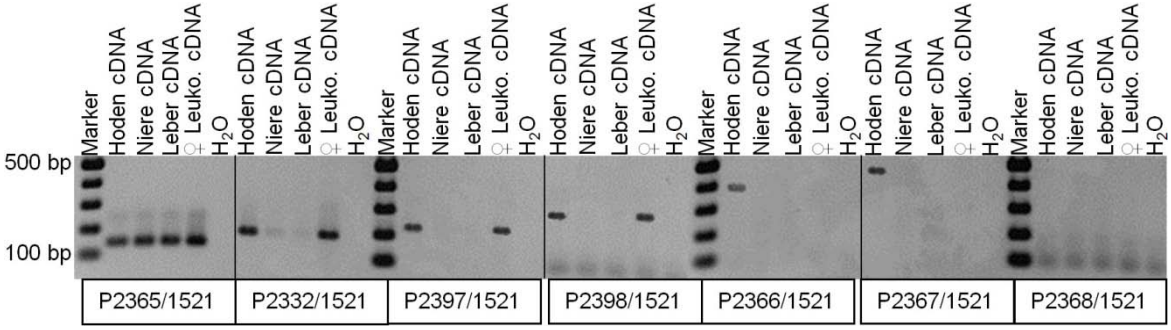


Abb. 4-44: RT-PCR Analyse Rind *Ddx3x*.
Errechnete PCR-Amplifikatgrößen für P2365/1521 = 144 bp, P2332/1521 = 190 bp, P2397/1521 = 221 bp, P2398/1521 = 277 bp, P2366/1521 = 391 bp, P2367/1521 = 436 bp, P2368/1521 = 489 bp. Alle somatischen Gewebe sind weiblichen Ursprungs. 35 PCR-Zyklen. DNA-Marker ist eine 100bp DNA-Leiter.

Für die Rind *Ddx3x* RNA-Expression bedeuten diese Ergebnisse also, daß zwischen den P2398 und P2366 Primerpositionen mind. eine weitere Initiationsstelle vorhanden sein sollte, die z. B. in Leukozyten genutzt wird. Diese Rind *Ddx3x* Transkripte sollten eine 5´UTR-Länge von mind. 218 bp haben (bis einschließlich P2398 Position berechnet). In dem fraglichen Sequenzabschnitt zwischen P2398 und P2366 sind mehrere Y-R Dinukleotide kartierbar. Ein mögliches Inr-Motiv ist in Abb. 4-41 markiert.

Darüber hinaus konnte eine hodenspezifische Rind *Ddx3x* 5´UTR-Verlängerung mit einer 5´UTR-Länge von 420 bp festgestellt werden. Für dieses Transkript wird die Initiationsstelle in Position 82.883 des genomischen Rind Contigs „004.81“ und die Bezeichnung „**Btau *Ddx3x* T-TSS-II“** vorgeschlagen. Diese Position ist homolog zur Human *DDX3X* T/B-TSS-II Initiationsstelle.

Das Vorhandensein einer verlängerten Rind *Ddx3x* 5´UTR-Sequenz wird durch ein *in-silico* vorhergesagtes *Ddx3x* Transkript („XM_864470.3“) unterstützt, das aus der NCBI-Datenbank extrahiert werden konnte. Dieses Rindertranskript beginnt in der Position 82.858 des genomischen Contigs „004.81“ (rev.-komplementär). Übertragen auf den Human BAC RP13-13A3 beginnt dieses Transkript in BAC-Position 30.338, 25 Nukleotide stromabwärts der Human *DDX3X* T/B-TSS-II Position. Mit dieser Analyse konnte also gezeigt werden, daß ein 5´ verlängertes Rind *Ddx3x* Transkript tatsächlich hodenspezifisch existiert. Zwischenzeitlich wurde der NCBI-Datenbankeintrag „XM_864470.3“ allerdings entfernt und durch „XM_864470.4“ ersetzt, der mit dem Eintrag „NM_001192962.1“ identisch ist und in der *Ddx3x* TSS-I Position beginnt.

Zusammenfassend für den *DDX3X/Ddx3x* Vergleich in Primaten und Säugern ist zu sagen, daß die TSS-I Region, mit kleinen Unterschieden in Maus und Rind, wahrscheinlich auch in anderen Säugetierspezies als ubiquitäre *DDX3X/Ddx3x* Initiationsstelle genutzt wird. Diese Core-Promoterregion ist somit bereits vor mindestens 100 Mio. Jahren in der Säugerevolution angelegt worden und ist seitdem in Sequenz und Funktion konserviert.
Als zweites Ergebnis ist die hodenspezifische *DDX3X/Ddx3x* Transkriptverlängerung zu nennen. Die dritte, hodenspezifische *DDX3X* Initiationsstelle T-TSS-III ist dagegen primaten-spezifisch. Obwohl die *DDX3X* T-TSS-III Core-Promotermotive im Sequenzalignment (Abb. 4-41) für das Rind erkennbar sind, wird diese Region erst in den Primaten funktionell genutzt. Das hodenspezifische *DDX3X* 5´UTR Splicing ist nach Auswertung aller RT-PCR Experimenten als humanspezifisch zu definieren.

Dieser speziesübergreifende Core-Promoter und RT-PCR Vergleich hat gezeigt, daß die Maus seit Trennung der verschiedenen Säugerlinien, im Vergleich zum Rind, stärkere Sequenz-veränderungen in den Core-Promoterregionen aufweist. Dies gilt für Maus *Ddx3x*, aber auch in besonderem Maße für Maus *Ddx3y* wie in Kapitel 4.4.2 dargestellt. Die Ergebnisse zeigen somit deutlich, daß, auch für *DDX3X*, nicht die Maus, sondern das Rind das besser geeignetere Versuchstier, für die Analyse der konservierten Kontrollelemente der *DDX3Y* und *DDX3X* Keimbahnfunktion in Mensch, darstellt.

4.6. Kartierung der potentiell funktionellen Promoterdomänen von Human *DDX3Y* und *DDX3X*

Nach Kenntnis der *DDX3Y* und *DDX3X* TSS- und Core-Promoterregionen wurde versucht, die dazugehörigen funktionellen Promoterdomänen darzustellen und einzugrenzen. Dazu werden *in-silico* die genomischen Sequenzen von *DDX3Y* und *DDX3X* bis 2,5 kb stromaufwärts der konservierten Translationsstartstellen (ATG) in Exon1 verglichen. Zur 3´ Justierung eines optimalen Alignments wird die hochkonservierte, kodierende Sequenz bis zum 3´Ende von Exon1 in das Alignment einbezogen. Zur Erkennung funktioneller Abschnitte in den Promoterregionen erfolgt der Vergleich auf konservierte Sequenzabschnitte, sowohl horizontal in den Human Genen *DDX3Y*[35] und *DDX3X*[36], als auch vertikal in Bezug zu den orthologen *DDX3Y* und *DDX3X* Genen in fünf anderen Mammalia (Schimpanse, Rhesusaffe, Weiß-büschelaffe, Rind, Maus). Hierdurch werden die spezifischen Sequenzveränderungen der Human *DDX3Y* und *DDX3X* Promoterregionen, seit der divergenten Entwicklung der Säuger-

[35] *DDX3Y*: BAC-Klon 475I1, Pos. 52.646 - 55.190, 2.545 bp.
[36] *DDX3X*: BAC-Klon RP13-13A3, Pos. 28.230 - 30.774, 2.545 bp.

Geschlechtschromosomen vor ~ 130 Mio. Jahren (Veyrunes et al., 2008), markiert. Die Evolution der *DDX3Y* und *DDX3X* Promoterregionen in den verschiedenen Säugerspezies wird somit sichtbar gemacht.

Ausgangspunkt ist der *in-silico* ein Sequenzvergleich, mit dem Programm mVISTA (Frazer et al., 2004), mit dem evolutionär konservierte Sequenzabschnitte, sogenannte „conserved non-coding sequences“ (CNS) markiert werden. CNS funktionieren häufig als *cis*-regulative Elemente, da in ihnen konservierte Transkriptionsfaktor-Bindestellen (cTFBS) lokalisiert sind (Hardison 2000; Loots et al., 2000; Elgar & Vavouri, 2008; Loots, 2008). In einem zweiten Schritt werden die potentiellen Human *DDX3Y* und *DDX3X* Promoterregionen auf besondere DNA-Signaturen, wie z.B. homopolymere und binäre Nukleotidabschnitte, sowie potentielle CpG-Islands (CGIs), untersucht. Im dritten Schritt werden potentiell funktionelle TFBSs kartiert. Durch den horizontalen und vertikalen Vergleich der TFBS-Karten, wird schließlich aus diesen Karten das konservierte TFBS-Muster herausdestilliert. Corà et al. (2004, 2005) zeigten, daß für die Genregulation funktionell wichtige TFBS in den stromaufwärts gelegenen Promoterregionen meist signifikant häufiger vorkommen, als es die normale statistische Verteilung von TFBS voraussagen würde. Darüber hinausgehend werden sogenannte TF-Module, Einheiten synergistisch agierender TFs, die wohl wichtig für die korrekte Transkriptionsregulation sind (Istrail & Davidson, 2005; Smith et al., 2006), identifiziert und kartiert. Ebenso können bestimmte TFBS-Abfolgen funktionell wichtig sein, wenn in zwei unterschiedlichen Promoterregionen zwei oder mehr TFBSs in konservierten Abständen zueinander lokalisiert sind. Zusätzlich erfolgt daher die Kartierung über-repräsentierter TFBSs, auffällig repetitiver TFBSs, sowie eine Analyse der Abstände zwischen den cTFBS auf Konserviertheit. All diese *in-silico* Daten tragen dazu bei, die in Kapitel 4.9 folgende experimentelle Promoteranalyse, auf Promoterbereiche zu fokussieren, die mit hoher Wahrscheinlichkeit tatsächlich funktionell sind.

4.6.1. VISTA-Plot zur Analyse konservierter Bereiche in den Human *DDX3Y* und *DDX3X* Promoterregionen

In Abb. 4-45 ist der mVISTA-Plot der Human *DDX3Y* und *DDX3X* Promoterregion zu sehen, wobei die *MSY2* Repeateinheiten 2, 3 und 4 vor *DDX3Y*, zur Verbesserung der Alignment-qualität, entfernt wurden. mVISTA nutzt für die Sequenzalignments das Programm LAGAN (Limited Area Global Alignment of Nucleotides) für paarweise Alignments und MLAGAN (Multi-LAGAN) für multiple Alignments (Brudno et al., 2003; siehe Kapitel 3).
Der Vergleich der Human *DDX3Y* und *DDX3X* Promotersequenzen ergab eine X-Y Gesamthomologie von 50,1% (1.298 homologe nts von 2.545 nts). Die standardmäßig eingestellten mVISTA CNS-Parameter entsprechen der Definition von Loots et al. (2000), die

eine ununterbrochene genomische Sequenz von mindestens 100 Nukleotiden mit einer Homologie von 70% oder höher fordert. Dermitzakis et al. (2005) merken allerdings an, daß die Loots CNS-Kriterien sehr stringent gewählt sind und funktionell konservierte *cis*-Elemente durchaus geringere Homologiewerte aufweisen können. Um konservierte Sequenzabschnitte zwischen der Human *DDX3X* und *DDX3Y* Promotersequenzen zu identifizieren, wurden in dieser Arbeit die mVISTA Parameter auf 30 bp Länge und 50% Homologie gesetzt (Abb. 4-47, A-1 bis A-3). Ohne Anpassung der mVISTA Parameter würde nur ein Sequenzabschnitt im X-Y konservierten, kodierenden Bereich von Exon1 den stringenten Kriterien entsprechen. Dermitzakis et al. (2005), sowie Rödelsperger et al. (2009), mit ihrer Definition einer „short ultraconserved promoter region“ (SUPR; 30 bp Länge, 98% Sequenzhomologie in sieben Säugerspezies, darunter Opossum und Mensch) zeigen, daß je nach analysierter Sequenz eine Anpassung der CNS-Kriterien notwendig sein kann, um auch kurze Sequenzelemente, wie TFBS und Core-Promotermotive, identifizieren zu können, die mit den stringenten Parametern nach Loots et al. (2000) häufig nicht erfasst werden.

Angewandt auf die potentiellen Promoterregionen von Human *DDX3Y* und *DDX3X* sind neun konservierte Regionen mit Sequenzhomologien über 50% zu unterscheiden (Abb. 4-45, Tab. 4-7). Sie werden in dieser Arbeit als „Konservierte Region in Human“ (KRiH) bezeichnet. Die konservierte Region, welche im Bereich des Exon1 lokalisiert ist, entspricht dabei als einzige Region den stringenten CNS-Kriterien und wird daher als „CNSiH-1“ bezeichnet werden. Da in der CNSiH-1 auch nicht-kodierende Sequenzen vor dem ATG aus Exon1 zu finden sind und *cis*-regulative Elemente auch stromabwärts einer TSS lokalisiert sein können, wird CNSiH-1 als eigenständige Region aufgeführt. Aus Abb. 4-45 und Tab. 4-7 ist zu sehen, daß die Sequenzregionen der kartierten Keimzell-spezifischen TSS der beiden Gene *DDX3Y* (T-TSS-II, T-TSS-I) und *DDX3X* (T-TSS-III) außerhalb der X-Y homologen Regionen lokalisiert liegen. Es handelt sich also jeweils um Allel-spezifische Startstellen, die sich erst nach der Trennung der Geschlechtschromosomen entwickelt haben. Die KRiHs-1 bis -3 liegen stromaufwärts zu *MSY2-1* bzw. *MSY2-X*. Innerhalb von KRiH-1 und KRiH-2 konnte je eine hochkonservierte Region identifiziert werden, die über die kürzere Strecke von 50 bp eine hohe Sequenzhomologie von 86% (KRiH-1 *DDX3Y*: Pos. 356 - 405; *DDX3X*: Pos. 493 - 544) bzw. von 72,5% (KRiH-2 *DDX3Y*: Pos. 741 - 792; *DDX3X*: Pos. 851 - 901) aufweist.

KRiH-4 beginnt vor *MSY2-1* bzw. *MSY2-X* und endet jeweils mittig in diesen. KRiH-5 befindet sich kurz hinter der *DDX3Y* T-TSS-II bzw. kurz vor der *DDX3X* T-TSS-III Initiationsstelle. KRiH-6 befindet sich innerhalb von *DDX3Y* Exon-Ta und am Anfang von Exon-Tb. In der *DDX3X* Sequenz befindet sich dieser Block zwischen den T-TSS-III und T/B-TSS-II Initiationsstellen. Die Exon-Ta 5´-GT Splicestelle liegt in der KRiH-6, ist aber selbst nicht konserviert. Die *DDX3Y* 3´-AG-Splicestellen sind in KRiH-8 bzw. CNSiH-1 lokalisiert. Die *DDX3X* 5´-GC-Splicestelle

befindet sich zwischen KRiH-6 und KRiH-7 in einem wenig konservierten Bereich. Die *DDX3X* 3´-AG Splicestelle ist in KRiH-8 lokalisiert.

Tab. 4-7: Liste der mit mVISTA ermittelten konservierten Regionen (KRiH) zwischen der *DDX3Y* und *DDX3X* Promoterregion.

Spalte 2: Position der ermittelten Elemente im Y-Chr. BAC 475I1, bezogen auf die verkürzte Sequenz (ohne *MSY2-2/3/4*). In Klammern sind die Positionen der KRiHs für die vollständige, 2,55 kb lange Human *DDX3Y* Promotersequenz angegeben. Spalte 3: Position der ermittelten Elemente in der 2.237 bp langen *DDX3Y* Sequenz (ohne *MSY2-2/3/4*). Spalte 4: Länge der ermittelten Elemente in *DDX3Y* Promoterregion. Spalte 5: Position der ermittelten Elemente im X-Chr. BAC RP13-13A3. Spalte 6: Position der ermittelten Elemente in der 2.545 bp langen *DDX3X* Sequenz. Spalte 7: Länge der ermittelten Elemente in *DDX3X* Promoterregion. Spalte 8 gibt die Sequenzidentität der konservierten Elemente in % an. Die Berechnung der Sequenzhomologie erfolgt nach folgender Formel: [$Anzahl\ identischer\ Nukleotide \times 100\% \div Nukleotide\ kürzere\ Sequenz = \%\ Homologie$]. Werte in Klammern geben die Homologiewerte gegen die jeweils längere Sequenz an, da mVISTA nicht generell gegen die kürzere Sequenz berechnet, sondern jeweils nur gegen eine bestimmte Sequenz. KRiH = „Konservierte Region in Human“. CNS ist wie in Abb. 4-45 rosa markiert.

Homologe Sequenz-blöcke	*DDX3Y*			*DDX3X*			% Homologie
	BAC-Position	Position 2,24 kb Sequenz	bp Länge	BAC-Position	Position 2,55 kb Sequenz	bp Länge	
KRiH-1	52.859 - 53.060	214 – 415	202	28.559 - 28.783	330 – 554	225	68,3 (61,3)
KRiH-2	53.321 - 53.551	676 – 906	231	28.994 - 29.257	765 – 1.028	264	65,8 (57,6)
KRiH-3	53.632 - 53.717	987 – 1.072	86	29.333 - 29.432	1.104 – 1.203	100	61,6 (53,0)
KRiH-4	53.801 - 53.909	1.156 – 1.264	109	29.515 - 29.629	1.286 – 1.400	115	56,9 (53,9)
KRiH-5	54.351 - 54.383	1.398 – 1.430 (1.706 – 1.738)	33	29.766 - 29.796	1.537 – 1.567	31	51,5 (56,7)
KRiH-6	54.474 - 54.667	1.521 – 1.714 (1.829 – 2.022)	194	29.951 - 30.142	1.722 – 1.913	192	58,3 (57,7)
KRiH-7	54.767 - 54.814	1.814 – 1.861 (2.122 – 2.169)	48	30.302 - 30.352	2.073 – 2.123	51	54,2 (51)
KRiH-8	54.864 - 55.096	1.911 – 2.143 (2.219 – 2.451)	233	30.452 - 30.670	2.223 – 2.441	219	56,7 (60,3)
CNSiH-1	55.097 - 55.190	2.144 – 2.237 (2.452 – 2.545)	94	30.671 - 30.774	2.442 – 2.545	104	78,7 (71,2)

Die KRiH-7 ist in *DDX3Y* Intron-T lokalisiert und beinhaltet die *DDX3X* T/B-TSS-II Sequenzregion. Der homologe Block KRiH-8 beginnt strom-aufwärts zu *DDX3Y* TSS-I-(T) und endet unmittelbar vor TSS-I(b). In der *DDX3X* Sequenz beginnt KRiH-8 zwischen T/B-TSS-II und der allgemeinen TSS-I und endet ebenfalls zwischen der ubiquitären TSS und dem ATG. Die CNSiH-1 befindet sich jeweils im Exon1. Da KRiH-8 und CNSiH-1 direkt ineinander übergehen, können sie auch als KRiH-8b zusammengefaßt werden. KRiH-8b ergibt für *DDX3Y* einen Block von 327 nts mit 63% und für *DDX3X* 323 nts mit 63,8% X-Y Sequenzhomologie.

Bereits bekannte Enhancer-Elemente konnten *in-silico* weder für *DDX3Y* noch für *DDX3X* in der jeweiligen Gen- bzw. Chromosomenregion identifiziert werden. Die Suche nach bereits

experimentell validierten, *cis*-aktiven Enhancer-Elementen wurde mit dem Programm „VISTA Enhancer Browser“ durchgeführt (Visel et al., 2007; Prabhakar et al., 2006). Grund für das negative Ergebnis war die ausschließliche Verwendung von „ultra-konservierten Elementen“ (UCR: 200 bp mit 100% Sequenzhomologie ohne Insertion oder Deletion; Bejerano et al., 2004) für den *in-vivo* Screen in transgenen Mäusen. Die UCRs stammen aus dem Vergleich drei verschiedener Säuger (Mensch, Maus, Ratte). Solche UCRs sind demnach nicht in der Nähe der *DDX3Y* und *DDX3X* Genloci lokalisiert.

Zur Klärung, ob die beobachtete Homologie von 50% zwischen den untersuchten Human *DDX3Y* und *DDX3X* Promotersequenzen höher oder niedriger ausfällt, als flankierende X-Y Sequenzabschnitte, wurde ein mVISTA-Plot des Y-spezifischen AZFa-Locus [*USP9Y–DDX3Y–UTY*] mit dem homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*], aus Stratum 3, durchgeführt. Die Sequenzen wurden so gewählt, daß je 5 kb 5´ von *USP9Y*/*USP9X* und 5 kb 3´ von *UTY*/*UTX* (*KDM6A*) die Gencluster begrenzen. Die Länge der Y-Sequenz beträgt 789,4 kb, die homologe X-Sequenz hat eine Länge von ~ 4,04 Mb. Der X-chromosomale Gencluster erstreckt sich somit über eine 5,11fach größeren Bereich, als im Y-Chromosom (19,55%). Dieser Wert ist auch nochmals geringer, als das Verhältnis der X- und Y-Chromosom Längen zueinander. Die Y-Chromosom Länge beträgt nur ~ 38,25% des X-Chromosoms. Mit den stringenten Parametern (mind. 70% Homologie, mind. 100 bp Länge) konnten in der untersuchten 789,4 kb langen Y-Sequenz insgesamt lediglich 444 X-Y konservierte Regionen (durchschnittlich ein Element pro 1.778 bp) identifiziert werden, was einem Anteil von 18,2% (143,6 kb, mit durchschnittlich 82,6% Homologie) der untersuchten Y-Sequenz entspricht (Abb. A-1). Auch entfallen bereits viele dieser 444 Sequenzelemente auf die ~ 100 konservierten Exone der Gene *USP9Y*, *DDX3Y*, *CASKP1* und *UTY* und in einem 10 kb Fenster stromaufwärts der *DDX3Y* Transkriptionseinheit sind weniger als zehn solcher konservierter Sequenzen lokalisiert, wobei keine Sequenz innerhalb der hier analysierten 2,5 kb Sequenz stromaufwärts des *DDX3Y* ATGs zu finden ist. Die Reduzierung der Parameterstringenz (50% Homologie, 30 bp Länge) führte zu 3.216 X-Y konservierten Regionen (durchschnittlich ein Element pro 245 bp) in der untersuchten 789,4 kb langen Y-Sequenz (Abb. A-3). Ein X-Y Sequenzvergleich, über eine vergleichbare Länge von 2,5 kb, von potentiell neutral oder nahezu neutral evolvierten intergenischen oder intronischen Sequenzen, im Bereich der AZFa Region, war auf Grund der starken Reduktion und Sequenzdivergenz der Y-chromosomalen Sequenz nicht möglich. Die Abb. A-1 bis A-3 zeigen deutlich, daß in den untersuchten 2,5 kb *DDX3X/DDX3Y* Promoterregionen die Sequenzhomologie auf ~ 50% absinkt. Dabei fällt aber auf, daß im Vergleich zu anderen Bereichen, ein Alignment von 2,5 kb nicht-kodierender Sequenz möglich ist.

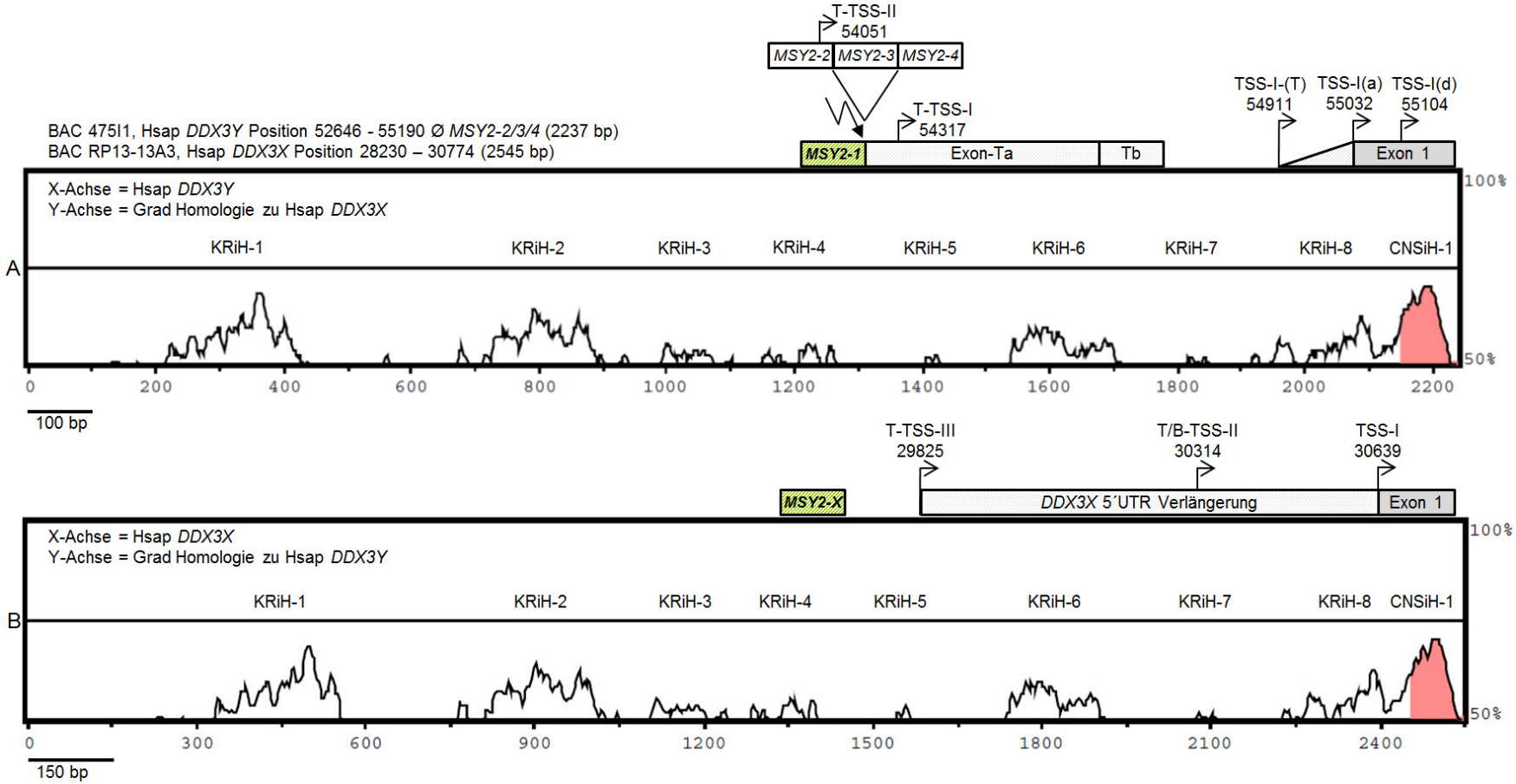


Abb. 4-45: Graphischer VISTA-Plot des Alignments zwischen *Hsap DDX3Y* und *DDX3X*.

A) Basislinie (X-Achse) ist *DDX3Y* (BAC 475I1, Position 52.646 – 55.190) ohne die *MSY2* Repeateinheiten 2, 3 und 4 (308 bp entfernt, Bruch durch Blitz angezeigt und als Einschub oberhalb gezeigt). Oberhalb des Plots ist die *DDX3Y* 5´UTR bis einschließlich Exon1 schematisch dargestellt, wie in Kapitel 4.1 (Abb. 4-8) erarbeitet. *DDX3Y* ATG an 55.146 (hier 2.193).
B) Basislinie (X-Achse) ist *DDX3X* (BAC RP13-13A3, Position 28.230 – 30.774). Oberhalb des Plots ist die *DDX3X* 5´UTR bis einschließlich Exon1 schematisch dargestellt, wie in Kapitel 4.2 (Abb. 4-13) erarbeitet. *DDX3X* ATG an 30.730 (hier 2.501). KRiH = „Konservierte Region in Human“. „Conserved non-coding sequences“ CNS sind rosa dargestellt. *MSY2-1* und *MSY2-X* sind grün schraffiert. Positionen in den schematischen Gendarstellungen können auf Grund des veränderten Maßstabes für *DDX3Y* nicht exakt verglichen werden.

4.6.2. Analyse der *DDX3Y* und *DDX3X* Promoterregionen auf DNA-Signaturen mit binärer und homopolymerer Nukleotidkomposition

Da bekannt ist, daß die Basenzusammensetzung eines DNA-Stranges starken Einfluß auf die Struktur, Windung und Krümmung der DNA, sowie auf die Organisation und den Besatz mit Histonen in den Nukleosomen nimmt (Schones et al., 2008; Radman-Livaja & Rando, 2010; Cairns, 2009), werden in diesem Abschnitt die potentiellen *DDX3Y* und *DDX3X* Promoterregionen auf das Vorhandensein binärer und homopolymerer DNA-Signaturen analysiert. Hervorgehoben wird dabei das Muster von CpG-Dinukleotiden als bereits bekanntes Promoter-Funktionselement in einem CpG-Island (CGI).

Für die Suche nach auffälligen homopolymeren und binären DNA-Signaturen in den potentiellen Promoterregionen wurde das Programm „TRACTS“ von Gal et al. (2003) verwendet. Als Signifikanz-Parameter für homopolymere Sequenzen wird eine Traktlänge von mindestens 10 Nukleotiden (nts) und für binäre Sequenzen eine Traktlänge von mind. 15 nts. Für beide Einstellungen gilt als Toleranz maximal ein Mismatch pro 10 Nukleotide. Funktionelle Kriterien für die Entfernung eines solchen Traktes zu einer TSS existieren nicht, sondern sind von der jeweiligen Promotersequenz und der Nukleosomen-Organisation abhängig (persönliche Mitteilung E. Sagal, 2010). Generell gilt aber, daß kurze Sequenzen, die den Einbau von Nukleosomen verhindern können, sogenannte „Nucleosome Excluding Sequences“ (NES), innerhalb der „Nucleosome-depleted Region“ (NDR) auftreten können. NDRs sind ~ 150 bis 200 bp stromaufwärts der TSS lokalisiert (Cairns, 2009; siehe Einleitung). Für diese Arbeit wird eine Entfernung von 150 nts (DNA-Länge um ein Histon-Oktamer) stromaufwärts zu einer TSS als relevante Entfernung für eine potentielle Funktion solcher Elemente angenommen.

Die Analyse ergab, daß von allen 23 kartierten homopolymären und binären Sequenzabschnitten für *DDX3Y* nur die zwei binären R-Trakt Nr. 5 und 6 vor TSS-I-(T) und der binäre K-Trakt Nr. 4 vor TSS-I(a) in direkter Nähe stromaufwärts einer TSS lokalisiert sind (Tab. 4-8). In G-reichen R- und K-Sequenzen können sogenannte „Quadruplex forming G-rich Sequences“ (QGRS) räumliche G-Quadruplex Strukturen entstehen, die Einfluß auf die Transkription nehmen können. Innerhalb des K-Traktes Nr. 4 ist ein TGG-Repeat vorhanden der das Potential für eine QGRS besitzt. Ein potentieller Einfluß der QGRSs auf die Translation wird in Kapitel 4.11.2 behandelt.

Für *DDX3X* konnten vor T-TSS-III und vor T/B-TSS-II binäre Nukleotidblöcke identifiziert werden. Vor *DDX3X* T-TSS-III geht eine (C,G)-reiche Region (S-Trakte Nr. 3 - 5) in eine lange (C,T)-reiche Region über, die auch bis weit in die 5’UTR der T-TSS-III Transkripte hinein reicht.

Tab. 4-8: Auflistung aller homopolymeren und binären Sequenzabschnitte in den 2,5 kb Promoter-regionen von Human *DDX3Y* und *DDX3X*.

Parameter für homopolymere Trakte sind mind. 10 nts und max. 1 Mismatch/10 nts und für binäre Trakte mind. 15 Nukleotide und max. 1 Mismatch/10 nts. Grüne Trakte sind positionshomolog zwischen *DDX3Y* und *DDX3X*. Grau unterlegte Trakte sind im Bereich von 150 bp stromaufwärts zu einer TSS lokalisiert. Positionsangaben entsprechen der Lokalisation innerhalb der 2,5 kb Sequenz. *DDX3Y* = BAC 475I1 (Position 52.646 - 55.190), *DDX3X* = BAC RP13-13A3 (Position 28.230 - 30.774). Ø = kein, --- = nicht vorhanden.

	DDX3Y				***DDX3X***			
	Position	**Länge**	**Mismatch**	**Lage & Abstand zu TSS**	**Position**	**Länge**	**Mismatch**	**Lage & Abstand zu TSS**
A-Trakt	132 - 142	11	1	---	1960 - 1984	25	2	126 bp vor T/B-TSS-II
C-Trakt	---	---	---	---	1720 - 1729	10	1	---
G-Trakt	---	---	---	---	1452 - 1463	12	1	145 bp vor T-TSS-III
T-Trakt	1) 196 - 205 2) 909 - 935	1) 10 2) 27	1) 1 2) 2	---	---	---	---	---
R-Trakt (A,G)	1) 486 - 528 2) 2030 - 2053 3) 2057 - 2078 4) 2080 - 2099 5) 2136 - 2155 6) 2189 - 2203	1) 43 2) 24 3) 22 4) 20 5) 20 6) 15	1) 4 2) 2 3) 2 4) 2 5) 2 6) 1	5) 131 bp vor TSS-I-(T) 6) 78 bp vor TSS-I-(T)	1) 814 - 835 2) 947 - 969 3) 1920 - 1942 4) 1957 - 1997 5) 2014 - 2028 6) 2040 - 2058	1) 22 2) 23 3) 23 4) 41 5) 15 6) 19	1) 2 2) 2 3) 2 4) 4 5) 1 6) 1	4) 129 bp vor T/B-TSS-II 5) 72 bp vor T/B-TSS-II 6) 46 bp vor T/B-TSS-II
Y-Trakt (C,T)	1) 196 - 211 2) 903 - 935 3) 2481 - 2496	1) 16 2) 33 3) 16	1) 1 2) 3 3) 1	---	1) 1133 - 1147 2) 1587 - 1605 3) 1626 - 1749	1) 15 2) 19 3) 124	1) 1 2) 1 3) 12	2) 10 bp vor T-TSS-III
K-Trakt (G,T)	1) 191 - 205 2) 907 - 936 3) 1063 - 1077 4) 2306 - 2336	1) 15 2) 30 3) 15 4) 31	1) 1 2) 3 3) 1 4) 3	4) 82 bp vor TSS-I(a)	1) 586 - 601 2) 897 - 913	1) 16 2) 17	1) 1 2) Ø	1) --- 2) ---
M-Trakt (A, C)	122 - 164	43	4	---	1957 - 1996	40	4	129 bp vor T/B-TSS-II
S-Trakt (C,G)	---	---	---	---	1) 1274 - 1294 2) 1319 - 1341 3) 1379 - 1410 4) 1446 - 1480 5) 1522 - 1539 6) 1563 - 1592 7) 1600 - 1619	1) 21 2) 23 3) 32 4) 35 5) 18 6) 30 7) 20	1) 2 2) 2 3) 3 4) 3 5) 1 6) 3 7) 2	4) 151 bp vor T-TSS-III 5) 75 bp vor T-TSS-III 6) 34 bp vor T-TSS-III
W-Trakt (A,T)	1) 226 - 248 2) 385 - 440 3) 531 - 550 4) 588 - 607 5) 696 - 718 6) 901 - 932	1) 23 2) 56 3) 20 4) 20 5) 23 6) 32	1) 2 2) 5 3) 1 4) 2 5) 2 6) 3	1) --- 2) --- 3) --- 4) --- 5) --- 6) ---	1) 345 - 374 2) 512 - 551 3) 1960 - 1984	1) 30 2) 40 3) 25	1) 3 2) 4 3) 2	1) --- 2) --- 3) ---

Vor T/B-TSS-II ist eine (A,G)-reiche Region (R-Trakte Nr. 4 - 6) festzustellen. Homopolymere Sequenzen, die laut Literatur einen großen Einfluß auf die DNA-Windung haben, konnten vor *DDX3X* T-TSS-III (G-Trakt) und vor T/B-TSS-II (A-Trakt) lokalisiert werden. Es fällt auf, daß die binären und homopolymeren DNA-Signaturen bei *DDX3X* besonders vor den gewebe-spezifischen TSSs positioniert sind. Das gemeinsame Vorkommen von CpG-Island und einer

potentiellen NDR (hier A-Trakt) in der *DDX3X* Promoterregion, entspricht den Beobachtungen von Choi (2010), daß NDRs im Humangenom nur in CGIs vorkommen und zusammen eine aktive Chromatin-Konformation markieren.

Im Unterschied dazu, sind vor den *DDX3Y* T-TSS-II und T-TSS-I keine binären und homopolymeren Trakte zu finden und es scheint beinahe, als würden diese vermieden werden. Die homopolymeren T- und A-Trakte in der distalen *DDX3Y* Promoterregion liegen jeweils mehrere Hundert Basenpaare vor der T-TSS-II.
Aus dem Vergleich der Lokalisation der KRiHs geht hervor, daß die meisten Trakte in den kaum konservierten Regionen zu finden sind. Von 23 gefundenen Sequenztrakten für *DDX3Y* liegen nur 7 innerhalb einer KRiH und für *DDX3X* sind von 24 gefundenen Trakten 10 Stück in den KRiHs wiederzufinden. Lediglich in KRiH-1 konnten die W-Trakte 1 und 2 an homologen Positionen identifiziert werden. Alle anderen Trakte sind individuelle Entwicklungen von *DDX3Y* oder *DDX3X*. Ein auffälliger Unterschied in der *DDX3X* Promoterregion ist die 163 bp lange Pyrimidin-reiche Sequenz im Bereich von T-TSS-III (Pos. 29.816 - 29.978; Y-Trakte 2 und 3), die in der *DDX3Y* Promoterregion gänzlich fehlt (Abb. 4-46).
Der Y-Trakt ist für die *DDX3X* Promoteraktivität von Interesse, da ein als TC-Box (TCCCTCCCTCCC) bezeichnetes Sequenzelement als Initiationselement dienen kann (McKeon et al., 1990; Hoffman et al., 1990; Cameron et al., 1992; DiMartino et al., 1994; Dörflinger et al., 1999). An den BAC-Positionen 29.902 (TCCCTCCCTCTC) und 29.943 (TCCCCTCCCCTCCC) können zwei TC-Boxen mit leichten Abweichungen zur vorgeschlagenen TC-Box Sequenz festgestellt werden. Ein synergistischer Effekt von TC-Box und GC-Boxen (SP1 Bindestellen) auf die Promoteraktivität ist bekannt (McKeon et al., 1990).

CpG-Island Analyse:

Aus den Daten wird sehr deutlich, daß hinsichtlich des (G+C)-Gehaltes ein großer Unterschied zwischen beiden Promoterregionen besteht. Für die *DDX3Y* Promoterregion konnte, im Gegensatz zur *DDX3X* Promoterregion, kein S-Trakt identifiziert werden. Die unterschiedliche Nukleotidkomposition der Promoterregionen sollte daher auch Einfluß auf das Ergebnis der nachfolgenden CpG-Island Analyse haben.

Für die gesamte *DDX3Y* 2,5 kb Sequenz konnte ein (G+C)-Gehalt von 44,2% (A+T = 55,8%) festgestellt werden, was etwas höher als der durchschnittliche (G+C)-Gehalt von 41% des Humangenoms ist. Für die *DDX3X* Sequenz konnte ein wesentlich höherer (G+C)-Wert von 53,52% (A+T = 46,48%) ermittelt werden. Mit einem Wert von über 50% (G+C)-Gehalt gilt die *DDX3X* Promoterregion somit als (C,G)-reiche Sequenz (Bajic et al., 2006). Die *DDX3Y* Promoterregion stellt sich folglich als (A,T)-reiche Region dar.

Das Programm „CpG Island Searcher“ verwendet die stringenten CpG-Island Such-Kriterien von Takai & Jones (2002): mindestens 500 bp Länge, mind. 55% (G+C)-Gehalt und mind. einen Wert von 0,65 für das Verhältnis von beobachteten zu erwarteten CpG-Dinukleotiden (O/E CpG) bezogen auf den G+C-Gehalt des CGIs. Das Programm “CpGPlot” (EMBOSS) verwendet die relaxierten Parameter (mind. 200 bp Länge, mind. 50% (G+C)-Gehalt und mind. einen Wert von 0,60 für das Verhältnis von O/E CpG) nach Gardiner-Garden & Frommer (1987), welches für die Analyse kürzerer CpG-Islets (< 200 bp) von Interesse ist.

In Tabelle 4-9 sind die Ergebnisse beider CGI-Analysen dargestellt. Der geringere Gesamt-(G+C)-Gehalt der *DDX3Y* Region spiegelt sich in der Anzahl und Länge der kartierten CGIs wider. Mit „CpG Island Searcher“ konnte für *DDX3Y* kein CGI identifiziert werden, was mit der Eigenschaft des Programms zusammenhängt nur lange, durchgehende CGIs zu identifizieren. Für *DDX3X* konnte mit den stringenten „Takai-Jones Kriterien“ (2002) ein durchgehendes CGI von 1764 bp, mit einem durchschnittlichen (G+C)-Gehalt von knapp 60%, kartiert werden. Mit 135 CpG-Dinukleotiden innerhalb von 1764 bp kommt diese Region auf eine O/E CpG-Frequenz von 0,853 und somit auf eine wesentlich höhere CpG-Häufigkeit (~1 CpG/100 nts), als für durchschnittliche genomische Sequenzen. Obwohl das CGI in der X-Y Homologieregion KRiH-2 beginnt, ist in der *DDX3Y* Promoterregion keine vergleichbare CpG-Dinukleotid-häufigkeit über die gesamte Sequenz feststellbar.

Tab. 4-9: Vergleich der pot. CpG-Islands in den Promoterregionen von Human *DDX3Y* und *DDX3X*.
Für *DDX3Y* = 2.545 bp, BAC 475I1 (Position 52.646 - 55.190). Für *DDX3X* = 2.545 bp, BAC RP13-13A3 (Position 28.230 - 30.774). „CpG Island Searcher“ ist so konfiguriert, daß benachbarte CGIs mit einer Lücke ≤ 100 bp vereint werden. Werte in Klammern in Spalte 5 und 9 (O/E CpG) geben die Anzahl der gefundenen CpG-Dinukleotide an.

	CpG Island Searcher (500 bp, 55%, 0,65 O/E CpG)				**EMBOSS CpGPlot** (200 bp, 50%, 0,6 O/E CpG)			
	BAC-Position	**Position (bp Länge)**	**%GC**	**O/E CpG**	**BAC-Position**	**Position (bp Länge)**	**%GC**	**O/E CpG**
DDX3Y	---	---	---	---	54.795 - 55.131	2.150 – 2.486 (337 bp)	51,04	1,25 (25 CpG)
DDX3X	29.011 - 30.774	782 – 2.545 (1.764 bp)	59,9	0,853 (135 CpG)	1) 29.349 - 30.144 2) 30.467 - 30.717	1) 1.120 – 1.915 (796 bp) 2) 2.238 – 2.488 (251 bp)	1) 67,46 2) 60,96	1) 0,82 (71 CpG) 2) 1,10 (25 CpG)

Die CGI-Analyse mit dem zweiten Programm „EMBOSS CpGPlot“ ergab für *DDX3Y* ein CGI von 337 bp (Pos. 54.795 – 55.131), welches in *DDX3Y* Intron T, 73 nts stromabwärts der Exon-Tb Splicestelle, beginnt und bis unmittelbar vor das ATG (Pos. 55.146) reicht. Mit 51,04%

(G+C)-Gehalt liegt es am unteren Limit zur CGI-Erkennung, hat aber durch 25 CpG-Dinukleotide in diesem Sequenzabschnitt einen hohen Wert von 1,25 für die O/E Häufigkeit von CpG-Dinukleotiden.
Für *DDX3X* konnten mit „CpGPlot“ zwei CGIs identifiziert werden. Das erste CGI von 796 bp Länge beginnt ~220 bp stromaufwärts zu *MSY2-X* und endet zwischen *DDX3X* T-TSS-III und T/B-TSS-II. Mit einem G+C Gehalt von 67,5% entspricht es sehr genau dem Durchschnittswert für CGIs im Humangenom. Die Zahl von 71 CpG-Dinukleotide (O/E CpG = 0,82) ist wesentlich höher, als für eine genomische Sequenz dieser Länge erwartet würde. Das 251 bp lange zweite CGI in der *DDX3X* Promoterregion beginnt zwischen den T/B-TSS-II und TSS-I Startstellen und endet, wie bei *DDX3Y*, unmittelbar vor dem ATG (Pos. 30.730). Der G+C Gehalt ist mit knapp 61% niedriger als beim ersten CGI, hat aber mit 25 lokalisierten CpG-Dinukleotiden eine höhere O/E CpG Häufigkeit von 1,1.
Der Vergleich der CGIs, in den proximalen Promoterregionen von *DDX3Y* und *DDX3X*, zeigt, daß von den jeweils 25 kartierten CpG-Dinukleotiden zehn an homologen Positionen lokalisiert sind.

Zur Klärung der Frage, ob in der *DDX3Y* und *DDX3X* Promoterevolution entweder ein CGI in der *DDX3Y* Promoterregion verloren gegangen ist, oder ob in der *DDX3X* Promotersequenz ein CpG-Island neu entstanden ist, wurden die Ergebnisse der Human *DDX3Y*/*DDX3X* CGI-Analyse mit der potentiellen Promotersequenz des orthologen *Ddx3* Gens aus *Monodelphis domestica* (Mdom), der Spitzmausbeutelratte, verglichen. Die in Mittel- und Südamerika vorkommende *Monodelphis domestica* gehört zur Unterklasse der Beutelsäuger (Marsupialia, Metatheria). Im Gegensatz zu den höheren Säugetieren (Eutheria) befindet sich das *Ddx3* Gen noch auf einem Autosom, hier auf Chromosom 4. Der Vergleich von Metatheria und Eutheria Genomen hat bereits gezeigt, daß die in Eutheria konservierten Protein-kodierenden Sequenzen nur einer geringen Veränderung im *Monodelphis* Genom unterlagen. Andererseits könnten bis zu 20% der in Eutheria konservierten CNS im *Monodelphis* Genom fehlen oder verändert sein (Lemos, 2007; Mikkelsen et al., 2007).
Für die Analyse wurde die genomische Sequenz von *Monodelphis domestica* Chromosom 4[37] im Bereich zwischen 22,34 und 22,33 Mb verwendet. Als Transkript stand eine 5´ unvollständige *Ddx3* cDNA aus der NCBI-Datenbank (XM_001378149.1) zur Verfügung. Im Bereich zwischen Mdom Exon1 bis incl. Intron2 besteht eine größere Sequenzierungslücke. Das genomische Alignment ist somit nur bis zur Position von Human *DDX3X* TSS-I und ab Exon3 möglich. Für Mdom wurde eine 3.089 bp lange Sequenz ermittelt, die homolog zu der Hsap *DDX3X* 2,5 kb Sequenz ist. Der Längenunterschied beruht auf indel-Ereignissen. Homolog zur Human *DDX3X MSY2-X* Sequenz konnte eine 129 bp lange Sequenz in der

[37] Referenzsequenz NW_001581956.1 revers-komplementär, Genom-Assembly MonDom5.

potentiellen Monodelphis Promoterregion identifiziert werden. Die Sequenz-homologie zur Human *MSY2-X* Sequenz beträgt 66,4%.
Für die gesamte Mdom Sequenz konnte ein (G+C)-Gehalt von 47,69% (A+T = 52,31%) ermittelt werden. Mit „CpG-Island Searcher“ konnte ein 1.316 bp langes, durchgehendes CGI mit 61,3% (G+C)-Gehalt und einer O/E CpG von 0,701 identifiziert werden. Das Mdom CGI beginnt 240 Nukleotide vor der homologen *MSY2* Sequenz. Die Analyse mit „CpGPlot“ ergab drei kleinere CGIs mit Längen zwischen von 331 bp (CpG-P1 = 66,47%, O/E CpG = 0,85), 228 bp (CPG-P2 = 63,16%, O/E CpG = 0,71) und 404 bp (CpG-P3 = 57,92%, O/E CpG = 0,93).

Das Vorhandensein eines langen CGIs in der potentiellen Mdom *Ddx3* Promoterregion entspricht daher im Aufbau eher der Human *DDX3X*, als der *DDX3Y* Promoterregion. Obwohl beide *DDX3X/Ddx3* Promoterregionen seit dem LCA ihre eigene Promoterevolution durchliefen, blieb die CGI-Domäne im *DDX3X* Promoter konserviert. In der Human *DDX3Y* Promoterregion ging somit ein langes CGI verloren.
Aus dem mVISTA-Vergleich zwischen der Mdom *Ddx3* und Hsap *DDX3X* Promoterdomäne geht auch hervor, daß bestimmte Promoterbereiche, wie *MSY2-X*, Y-Trakt und (C,G)-reiche Sequenzabschnitte bereits im Mdom *Ddx3* Promoter angelegt sind (Abb. 4-46). Der konservierte Bereich erstreckt sich somit von *MSY2-X* bis zu dem als KRiH-6 definierten Bereich vor der Human *DDX3X* T/B-TSS-II Position. Distal von *MSY2-X* sind keine Bereiche mit Homologien über 50% festzustellen.

Die Grundstruktur der *DDX3X* Promoterregion beginnt also bei *MSY-X* und besteht seit über 150 Millionen Jahren. Sie enthält vermutlich viele konservierte *cis*-Elemente, die als potentiell funktionelle Promoterelemente funktionieren.
Die *in-silico* Analysen der *DDX3Y* und *DDX3X* Promoterregionen machen deutlich, daß die Sequenzstrukturen der beiden Promoterregionen sehr divergent sind und sich insbesondere in der *DDX3Y* Promoterregion sehr spezifisch für das Y-Chromosom entwickelt hat. Die Zusammenfassung der CGI-Analyse und die Ergebnisse der DNA-Signatur-Analyse sind graphisch in Abb. 4-47 zu sehen.

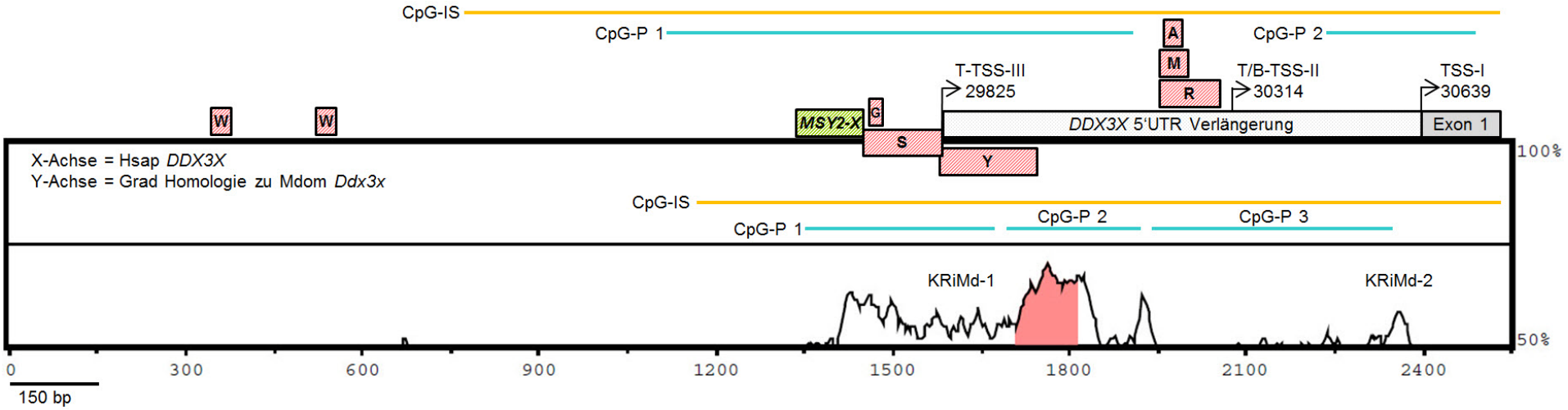


Abb. 4-46: Graphischer VISTA-Plot des Alignments zwischen Hsap *DDX3X* und Mdom *Ddx3*.

Basislinie (X-Achse) ist *DDX3X* (BAC RP13-13A3, Position 28.230 – 30.774). Oberhalb des Plots ist die *DDX3X* 5´UTR schematisch dargestellt, wie in Kapitel 4.2 (Abb. 4-13) erarbeitet. Die in Kapitel 4.6.2 ermittelten Sequenztrakte sind in Kastenform (rot schraffiert) angegeben. *DDX3X* ATG in BAC-Pos. 30.730 (hier 2.501). KRiMd = „Konservierte Region in *Monodelphis domestica*“. „Conserved non-coding sequences“. *MSY2-X* ist grün schraffiert. CNS ist rosa dargestellt.

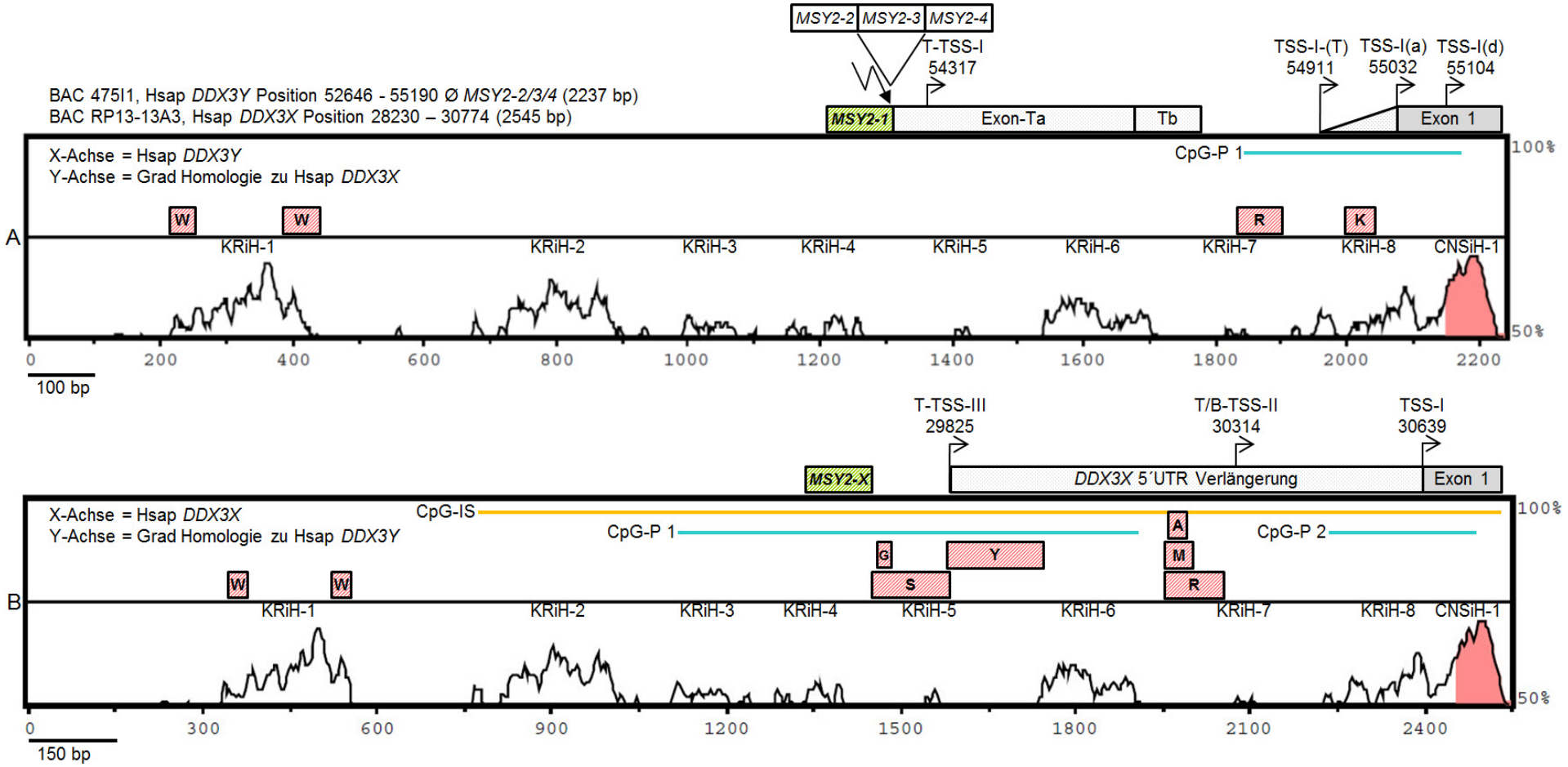


Abb. 4-47: Graphische Auswertung der CpG-Islands und DNA-Signaturen in den *DDX3Y* und *DDX3X* Promoterregionen.

Die Abbildung basiert auf Abb. 4-45. **A:** *DDX3Y* Promoterregion ohne die *MSY2* Repeateinheiten 2, 3 und 4 (308 bp entfernt, Bruch durch Blitz angezeigt und als Einschub oberhalb gezeigt). **B:** *DDX3X* Promoterregion. CpG-Islands mit „CpGPlot“ (CpG-P) sind hellblau, mit „CpG Island Searcher“ (CpG-IS) orange dargestellt. Die Boxen entsprechen den ermittelten Trakten aus Tab. 4-8. Angegeben sind die homologen W-Trakte und die Trakte, die 5´ im Abstand von max. 150 bp zu einer TSS lokalisiert sind. Die *DDX3Y* R-Trakte 5 und 6 sind zu einer Region zusammengefaßt. Die *DDX3X* R-Trakte 4 -6, die Y-Trakte 2 und 3 und die S-Trakte 4 bis 6 sind jeweils zu einer Region zusammengefaßt. *MSY2-1* und *MSY2-X* sind grün schraffiert. Die Sequenztrakte sind rot schraffiert. Die Trakt-Bezeichnungen richten sich nach den IUPAC Nomenklaturregeln.

4.6.3.1. Analyse konservierter Transkriptionsfaktorbindestellen und TFBS-Mustererkennung in den Human *DDX3Y* und *DDX3X* Promoterregionen

In diesem Abschnitt der Promoter-Analyse werden die potentiellen Transkriptionsfaktor-Bindestellen (TFBS) mit dem Genomatix-Programm „MatInspector“ (Cartharius et al., 2005) kartiert und mit „DiAlign professional TF“ auf X-Y Konserviertheit zwischen den Human Sequenzen untersucht. Die Hintergrundinformationen zum Verständnis der MatInspector-Analyse sind in Kapitel 2.14. beschrieben.

Für die proximalen und distalen Promoterregionen (in je 2.545 bp) von Human *DDX3Y* konnten mit MatInspector 658 potentielle TFBS (TF-Core ≥ 0,75; Optimized-Wert) und 643 potentielle TFBS für *DDX3X* kartiert werden. Komplette Listen aller *DDX3Y* und *DDX3X* TFBSs sind auf der beigefügten Daten-CD hinterlegt (Tab. A-16, A-17). Davon konnten 30 cTFBS mit dem Genomatix-Programm „DiAlign TF“ (Cartharius et al., 2005) selektioniert und kartiert werden, die auf 18 unabhängige Positionen, mit nicht überlappenden TF-Cores, reduziert werden können (Tab. 4-10, Abb. 4-48). Für 19 der cTFBS ist eine Expression in Hodengewebe der zugehörigen TFs bekannt. Informationen zur Expression der TFs in Hodengewebe wurden aus den verfügbaren Datenbanken von Genomatix, „Universal Protein Resource“ („UniProt“; Jain et al., 2009; The UniProt Consortium, 2010), „Human Protein Atlas“ („HPA“; Uhlén et al., 2005; Berglund et al., 2008) und aus der hier angegebenen Literatur ermittelt.

Für alle cTFBS konnte ein höherer Matrix-Wert als der definierte Optimized-Grenzwert festgestellt werden. Ein Matrix-Wert oberhalb von 0,85 konnte aber nicht für alle cTFBS festgestellt werden. Es gilt hier aber zu beachten, daß in dieser Analyse konservierte TFBS gesucht wurden und die oben formulierten Matrix-Selektionskriterien für Einzelsequenzanalysen konzipiert sind. Die cTFBS werden von „DiAlign TF“ zudem nicht nur für einen bestimmten TF, sondern immer für ganze TF-Familien angegeben. D.h., daß im *DDX3Y* Promoter eine TFBS für BRN2 und im *DDX3X* Promoter eine TFBS für BRN3 als homolog zueinander bezeichnet werden kann (siehe cTFBS Nr. 1, Tab. 4-10).

Der Vergleich der cTFBS Positionen mit der Lage der KRiHs zeigt eine deutliche TFBS-Akkumulation in KRiH-1 und -2. Für KRiH-1 konnten fünf nicht überlappende cTFBS und für KRiH-2 konnten vier cTFBS kartiert werden. Die (A,T)-reichen cTFBS BRN2/3, FHX, BRN2/3/5 und SATB1 liegen dabei in den konservierten W-Trakten in KRiH-1. Die drei cTFBS in KRiH-1, die zur TFBS Nr. 4 zusammengefaßt wurden (Tab. 4-10), sind im hochkonservierten Bereich von KRiH-1 lokalisiert. Die drei cTFBS in KRiH-2, die zur TFBS Nr. 7 zusammengefaßt wurden (Tab. 4-10), liegen dabei im hochkonservierten Bereich von KRiH-2 (Abb. 4-48).

Tab. 4-10: Konservierte TFBS (cTFBS) in Human *DDX3Y* und *DDX3X*.

Für *DDX3Y* = 2.545 bp, BAC 475I1 (Position 52.646 - 55.190). Für *DDX3X* = 2.545 bp, BAC RP13-13A3 (Position 28.230 - 30.774). Identische TF-Core- und Matrix-Werte zwischen Human *DDX3Y* und *DDX3X* werden nicht gesondert aufgeführt, veränderte Angaben für Human *DDX3X* sind in Klammern angegeben. In Spalte 2 können für *DDX3X* andere TF Bezeichnungen auftreten, da „DiAlign TF" cTFBS für TF-Familien angibt. In Spalte 3 sind die TFBS-Positionen entsprechend der Lage in der 2,5 kb Promoterregion angegeben. In Spalte 4 ist die Lage der TFBS in konservierter Region KRiH (Kapitel 4.6.1) angegeben. Spalte 5 gibt die Orientierung der TFBS an. „Core-Werte" in Spalte 6 geben Übereinstimmung zur TF-Core-Konsensussequenz an. Werte sind von 0 bis 1 angegeben, wobei 1 einer 100% Übereinstimmung zur Konsensussequenz entspricht. „Matrix-Werte" in Spalte 7 bewerten die Core-Sequenz und die umliegenden Nukleotide. „Opt-Werte" (Spalte 8) sind individuelle Grenzwerte für jede TF-Matrix und sollen nach Möglichkeit oberhalb von 0,85 liegen. Für die TFs BRN2 und OCT1 existieren jeweils mehrere TF-Matrizen in der BRN2 und OCT1 Matrix-Familie. Die individuellen TF-Matritzen einer Matrix-Familie unterschiedlichen sich im Opt-Wert und dem jeweiligen Erwartungswert einer Matrix (wie viele Bindungen an 1.000 bp zufälliger DNA-Sequenz). Als Matrix-Bibliothek wurde nur die Vertebraten Matrix-Gruppe gewählt. In Spalte 9 ist angegeben, ob für den TF eine Expression in Hodengewebe bekannt ist. Bei bekannter Hodenexpression ist Feld grün unterlegt. k. A. = keine Angabe zur Expression in Hodengewebe. Spalte 10 gibt die ermittelte Bindesequenz an. Die Sequenz kann revers-komplementär sein, wenn TF an Minus-Strang bindet. Für die exakten Algorithmen zur Berechnung der Matrizes siehe Quandt et al. (1995), Cartharius et al. (2005) und Kapitel 2.14.

Nr.	TF-Name in *DDX3Y* (*DDX3X*)	TF-Familie	Position	Lage KRiH	*DDX3Y*/ *DDX3X* Strang	Core-Wert	Matrix-Wert	Opt-Wert	Bekannte Hoden-expression	TF-Bindesequenz
1	BRN2 (BRN3)	POU TF-Familie	224 - 242 (350 - 368)	**1**	+	0,857 (1)	0,882 (0,807)	0,82 (0,78)	k. A.	actgtataaatTTATattg (AATaatttttaaataaaacag)
2	OCT1	POU TF-Familie	368 - 384 (506 - 522)	**1**	-	1	0,916 (0,924)	0,89 (0,8)	ja	ggCATGtaaatgaccgc (aaTATGtaaatgCcgcc)
3	FHX	„Forkhead domain" TF-Familie	379 - 395 (517 - 533)	**1**	+	0,818 (1)	0,831 (0,867)	0,83	ja	catgccAAAAtacatat (catattATAAtacattg)
4	BRN5	POU TF-Familie	385 - 407 (524 - 546)	**1**	+	1 (0,75)	0,843 (0,771)	0,74 (0,77)	k. A.	aaaataCATAttttattataaat (taatacattgtTTTAttataatc)
	BRN3 (BRN2)	POU TF-Familie	387 - 405 (525 - 543)	**1**	- (+)	0,75 (0,857)	0,809 (0,825)	0,78 (0,82)	k. A.	ttataataAAATAtgtatt (aatacattgttTTATtata)
	SATB1	CUT-Homeobox TF-Familie	390 - 404 (529 - 543)	**1**	-	1	0,961 (0,97)	0,94	ja	tatAATAaaatatgt (tatAATAaaacaatg)
5	FHX (XFD3)	„Forkhead domain" TF-Familie	416 - 432 (556 - 572)	**---**	+ (-)	1 (1)	0,833 (0,931)	0,83 (0,82)	ja	aattatATAAtaattag (gatctgtcAACAaacga)
	SRY	HMG-Box TFs	417 - 441 (533 - 557)	**--- (1)**	+	0,775 (1)	0,881 (0.985)	0,88	ja	attatATAAtaattaggtttgttag (gttttATTAtaatcattttgctttc)
6	BRN2	POU TF-Familie	706 - 724 (806 - 824)	**2**	+	0,933 (0,9)	0,883 (0,898)	0,86	k. A.	ttAATTttaaattcggaaa (tcGATTttaaaaattgaga)
7	OCT2	POU TF-Familie	757 - 773 (866 - 882)	**2**	-	0,775	0,833	0,8	ja	aaaATTCaaatacctcc (agaATTCaaatagttcg)
	OCT1	POU TF-Familie	760 - 776 (869 - 885)	**2**	+	1	0,847 (0,844)	0,81	ja	ggtatttgAATTttagc (actatttgAATTctggc)

7	CHR	"Cell cycle regulators: Cell cycle homology element"	761 - 773 (870 - 882)	**2**	+	1	0,986 (0,993)	0,92	k. A.	gtatTTGAatttt (ctatTTGAattct)
8	LHX3	LIM-Homeobox TF-Familie	805 - 827 (919- 941)	**2**	+	1	0,89 (0,923)	0,82	ja	aaatctggatattTAATctggtt (actgcgttttattTAATttcgta)
	HHEX	Homeobox TF-Familie	807 - 825 (921 - 939)	**2**	+	1	0,952 (0,977)	0,95	ja	atctggatattTAATctgg (tgcgttttattTAATttcg)
9	BRN3 (BRN2)	POU TF-Familie	808 - 826 (920 - 938)	**2**	- (+)	0,75 (1)	0,791 (0,932)	0,78 (0,92)	k. A.	accagattaAATAtccaga (ctgcgttttATTTaatttc)
8	TBP	„TATA-binding Protein" TF-Familie	811 - 827 (925 - 941)	**2**	+	0,75	0,79 (0,825)	0,78	ja	ggatattTAATctggtt (ttttattTAATttcgta)
10	DMRT 2 & 7 (DMRT1)	"Doublesex- and mab-3-related transcription TF"	1.285 - 1.305/ 1.287 - 1.307 (1.424 - 1.444)	**---**	-/+ (+)	0,75/1 (1)	0,76/0,94 (0,791)	0,75/0,81 (0,77)	ja	aaacaatgaAACAattatata/ tataATTGtttcattgtttcg (cgtgtctgtctcaTTGTtccc)
	SOX5	HMG-Box TFs	1.288 - 1.310 (1.425 - 1.447)	**---**	-	1	0,988 (0,982)	0,87	ja	tcacgaaaCAATgaaacaattat (tgagggaaCAATgagacagacac)
11	NRF1	NRF1/EWG TF-Familie	1.643 - 1.659 (1.465 - 1.481)	**---**	-	0,75 (1)	0,791 (0,899)	0,78	k. A.	ggtGTGCctgcgcttta (tgtGCGCgcgcgcctcc)
12	DEC2 (HES1)	bHLH TF-Familie	1.740 - 1.754 (1.569 - 1.583)	**---**	-	1 (0,944)	0,983 (0.929)	0,96 (0,92)	k. A.	gcagcaCGTGtgacc (gcggcgcGCGCaggt)
13	RXR/VDR	„NR2/1-Nuclear Hormone Receptor" TF-Familie	1.822- 1.846 (1.659 - 1.683)	**6 (---)**	-	1	0,887 (0,867)	0,85	ja	atcttcaggttaggGAGGtgaaaga (agagggagggaaaaGAGGggagcg)
14	YY1	YY TF-Familie	1.868 - 1.888 (1.763 - 1.783)	**6**	-	1	0,957 (0,978)	0,94	ja	agtgaCCATtttacgtgagcg (ggtggCCATtttgtgaagcag)
15	SPI-B (SPI1/PU.1)	ETS TF-Familie	2.033 - 2.053 (1.925 - 1.945)	**---**	+	1	0,915 (0,961)	0,88 (0,96)	ja	agtaaaagGGAAgagagagtg (ctgagagaGGAAgaagagtga)
16	HIF1-alpha/ARNT Heterodimer	bHLH TF-Familie	2.344 - 2.360 (2.335 - 2.351)	**7**	-	1	0,936 (0,913)	0,91	ja	acgcacaaCGTGcgtaa (acgtacaaCGTGtgccc)
17	MEL1	C2H2-Zinkfinger TF-Familie	2.493 - 2.509 (2.493 - 2.509)	**CNSiH-1**	+	1	0,999	0,99	k. A.	cttcaggGATGagtcat (cttcaggGATGagtcat)
	BACH1	bZIP-CNC TF-Familie	2.494 - 2.514 (2.494 - 2.514)	**CNSiH-1**	-	0,75	0,858	0,82	k. A.	accacaTGACtcatccctgaa (gccacaTGACtcatccctgaa)
	AP1	bZIP-Jun TF-Subfamilie	2.500 - 2.510 (2.500 - 2.510)	**CNSiH-1**	-	0,905	0,973	0,94	ja	catgACTCatc (catgACTCatc)
18	MESP1 (SCX)	TWIST-bHLH TF-Subfamilie	2.499 - 2.519 (2.500 - 2.520)	**CNSiH-1**	+	0,821 (1)	0,912 (0,934)	0,91	ja	tcaccacCACAtgactcatcc (gatgagtcatgTGGCagtgga)
	MITF	bHLH TF-Familie	2.502 - 2.516 (2.502 - 2.516)	**CNSiH-1**	+	1	0,95 (0,949)	0,92	k. A.	tgagtCATGtggtgg (tgagtCATGtggcag)
	c-MYC/MAX (USF1 & 2)	bHLH TF-Familie	2.504 - 2.516 (2.504 - 2.516)	**CNSiH-1**	-	0,86 (0,915)	0,927 (0,929)	0,92 (0,9)	ja	ccaccaCATGact (ctgcCACAtgact)

MSY2

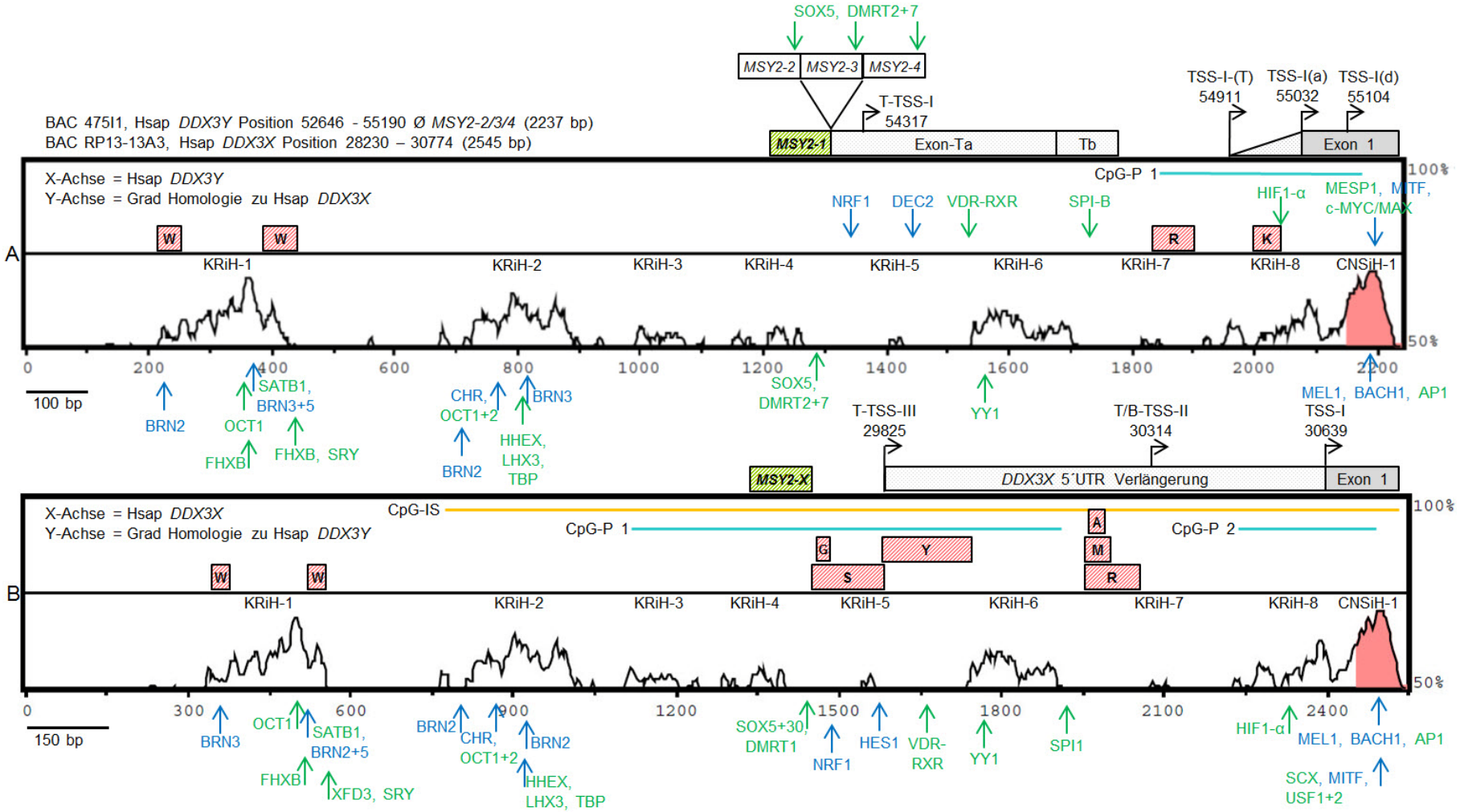


Abb. 4-48: Graphische Zusammenfassung der *DDX3Y* und *DDX3X* TF-Analyse.

Die Abbildung basiert auf Abb. 4-45. **A:** *DDX3Y* Promoterregion ohne die *MSY2* Repeateinheiten 2, 3 und 4 (308 bp entfernt, Bruch durch Blitz angezeigt und als Einschub oberhalb gezeigt). **B:** *DDX3X* Promoterregion. Angegeben werden die konservierten TFBSs in blau. Die cTFBSs mit bekannter Expression in Hodengewebe sind grün markiert. cTFBS-Positionen sind durch Pfeile dargestellt. Beschreibung der TFBSs siehe Tab. 4-10. *MSY2-1* und *MSY2-X* sind grün schraffiert. Die Sequenztrakte sind rot schraffiert.

Die Gesamtanzahl an konservierten und nicht-konservierten TFBSs beträgt für die *DDX3Y* KRiH-1 62 TFBSs und für die *DDX3X* KRiH-1 konnten 78 potentielle TFBS ermittelt werden. In KRiH-2Y konnten insgesamt 57 TFBS und in KRiH-2X 85 TFBS lokalisiert werden.

Entsprechend den verschiedenen Definitionen für Sequenzelemente mit Enhancer-Qualität, variiert die vorgeschlagene Anzahl an benötigten TFBS für ein funktionelles Enhancer-Element. Die Enhancer-Element Definition nach Levine & Tjian (2003) fordert mindestens zehn TFBSs von mindestens drei verschiedenen TFs. Für die Funktionalität des Interferon-β (IFN-β) „Enhanceosom" (dreidimensionaler Komplex aus Enhancer-Element und TFs) wurden acht TFs beschrieben (Merika & Thanos, 2001; Panne, 2008). Die X-Y konservierten Sequenzelemente KRiH-1 und -2 würden demnach, eine funktionelle Aktivität vorausgesetzt, diesen Definitionen für ein Enhancer-Element entsprechen.

In den KRiHs -3, -4, -5 und -7 konnten keine cTFBS festgestellt werden. In KRiH-6 ist für YY1 eine konservierte Bindestelle zu finden. In KRiH-8 ist die TFBS für den Faktor HIF1-α jeweils vor den ubiquitären TSS-I Positionen lokalisiert und in CNSiH-1 sind mehrere „basic Helix-Loop-Helix" (bHLH) Faktoren um den Translationsstartpunkt ATG gruppiert (siehe Tab. 4-10). Viele bHLH-TFs binden an eine E-Box mit dem Konsensus-Bindemotiv 5´-CANNTG-3´. Ob die Bindung an das Translationsstartcodon ATG eine funktionelle Bedeutung hat, oder die Folge einer zufällig vorhandenen Bindesequenz darstellt, ist nicht bekannt.

Die Bindestellen der Faktoren FHXB/XFD3, DMRT2&7/DMRT1, SOX5, NRF1, DEC2/HES1, SPI-B/SPI1 liegen außerhalb der konservierten Regionen. Hervorzuheben sind die Faktoren SOX5 und DMRT, die in dem konservierten Sequenzblock *MSY2/MSY2-X* lokalisiert sind.

Im Folgenden werden die TFs mit konservierten cTFBS und Matrix-Werten oberhalb von 0,85 und/oder bekannter Expression in Hodengewebe (siehe Tab. 4-10) detaillierter betrachtet.

OCT1

Der TF „Octamer-binding Protein 1" (OCT1 oder POU2F1) ist in KRiH-1 lokalisiert. OCT1, aus der POU TF-Familie, wird ubiquitär exprimiert, aber in Hodengewebe nur in Spermatogonien und Spermatozyten, nicht in Spermatiden (HPA; Sturm et al., 1988; Schöler et al., 1989; Phillips & Luisi, 2000). Aus der Maus sind von Oct1 in Hodengewebe verschiedene Proteinisoformen mit unterschiedlichen C-Termini bekannt (Riss & Laskov, 1999). Im Maus *Nanog* Promoter, einem entscheidenden Stammzell- und Keimzell-TF, wurde ein Octamer-binde Motiv identifiziert, welches wichtig für die Transkriptionskontrolle von *Nanog* ist. Diese TFBS kann von Oct1, aber auch von Oct4 und Oct6 erkannt werden (Wu & Yao, 2005).

FHX

Die zweite cTFBS in KRiH-1 gehört zu dem TF „Fork-head homologous X“ (FHX oder FOXJ2), der zwei unterschiedliche DNA-Motive erkennen kann und in zwei Proteinisoformen exprimiert wird (Pérez-Sánchez et al., 2000a, 2000b; Gómez-Ferrería & Rey-Campos, 2003). In der Maus wird Fhx u.a. in Keimzellen und in Sertoli-Zellen exprimiert (Granadino et al., 2000).

SATB1

Der ebenfalls in KRiH-1 lokalisierte Faktor SATB1, ist entscheidend bei der Organisation des Chromatins in Loop-Strukturen beteiligt. SATB1 gilt als ein spezifisches S/MAR-binde Protein („Scaffold/Matrix Attachment Region“), welches in der Lage ist, das Chromatin in viele kleine Transkriptions-aktive Schleifen zu falten (Cai et al., 2006; Göndör & Ohlsson, 2006). SATB1 bindet gewebespezifisch an AT-reiche Sequenzen, die dauerhaft ungepaart als Einzelstrang-DNA (BUR, „base-unpairing region“) vorliegen (Wang et al., 2010). Diese BURs sind in MARs (Matrix Attachment Regions) zu finden. SATB1 selbst dient als Plattform für weitere Chromatin-Remodelling Faktoren und Komplexe, die durch SATB1 an die S/MARs rekrutiert werden (Yasui et al., 2002; Cai et al., 2006; Galande et al., 2007; Kadauke & Blobel, 2009; Wang et al., 2010). Von SATB1 ist eine starke somatische Expression spezifisch in Thymozyten bekannt. Geringe Mengen SATB1 sind aber auch in Hodengewebe, Osteoblasten und in fötalem Hirngewebe zu finden (Dickinson et al., 1992; Nakagomi et al., 1994; Kohwi-Shigematsu et al., 1997; Alvarez et al., 2000; Hawkins et al., 2001).

SRY

Mit SRY („Sex-determining region Y protein“) wurde der lange gesuchte TDF („testis-determining factor“) identifiziert. SRY gilt daher als Gründungsmitglied der SOX-TF-Familie und wird in der A-Gruppe der SOX-Faktoren geführt. Die „High Mobility Group“-Proteindomäne (HMG) der SRY/SOX-Faktoren bindet ausschließlich in der kleinen Rinne („minor groove“) der DNA-Doppelhelix, was zur Folge hat, daß diese Faktoren die DNA biegen können und somit beim Remodelling der Chromatinstruktur in Promoterregionen beteiligt sind. Auf Grund ihrer Eigenschaften werden die HMG-Proteine als „Chromatin-assoziierte Nicht-Histon-Proteine“ oder „architektonische Transkriptionsfaktoren“ bezeichnet (Bustin & Reeves, 1996). TFs der SOX-Familie werden generell als Entwicklungsregulatoren beschrieben, die in vielen Entwicklungsprozessen Schlüsselpositionen einnehmen (Prior & Walter, 1996; Chew & Gallo, 2009; Wegner, 2010).

SRY ist entscheidend für die Maskulinisation der undifferenzierten Gonadenanlage während der Embryonalentwicklung (Kashimada & Koopman, 2010; Sekido, 2010). Die Hauptaufgabe von SRY besteht darin, die SOX9-Proteinexpression so zu erhöhen, das SOX9 die Sertoli-Zell-Differenzierung induzieren kann (Kanai et al., 2005; DiNapoli & Capel, 2008; Sekido & Lovell-

Badge, 2008; Barrionuevo & Scherer, 2010). Das SRY/SOX9-System bietet somit die Möglichkeit Entwicklungsprogramme in undifferenzierten Organanlagen zu studieren.

CHR

Das "cell cycle genes homology region" DNA-Motiv (CHR) in KRiH-2 ist eine TF-Bindesequenz, die häufig in Promotersequenzen von Genen, mit hoher Expression in der S- und G_2-Phase des Zellzykluses, vorkommt (Müller & Engeland, 2009). Die Gene, die unter Kontrolle des CHR-Motives stehen, werden in der G_0- und G_1-Phase reprimiert. D.h., daß die CHR-regulierten Gene während des Zellzyklus in Wellen exprimiert werden. Welche Faktoren an das CHR-Motiv binden können ist bisher nicht genau bekannt (Müller & Engeland, 2009). Die Expression CHR-regulierter Gene wird durch zwei bis drei CCAAT-Boxen aktiviert, an die NF-Y bindet. CHR-Elemente kommen häufig mit CDEs („cell cycle-dependent element") zusammen vor, die vier Nukleotide vor einem funktionellen CHR liegen. In der *DDX3Y* und *DDX3X* Promoterregion konnte kein CDE stromaufwärts zur CHR lokalisiert werden.

SOX5

Die in *MSY2*/*MSY2-X* konservierte TFBS für SOX5 (SRY-related HMG box) hat für beide Promoterregionen einen Matrix-Wert > von 0,98, trotz der langen Bindesequenz von 23 Nukleotiden. Der SOX5-TF, der zur D-Gruppe der SOX-Faktoren gehört, wird in drei somatischen Proteinisoformen, L-SOX5[38] genannt, und in einer kürzeren hodenspezifischen Variante[39] exprimiert. Dies führt zu vier unterschiedlichen Proteinisoformen (L-SOX5: 763 AS, 750 AS, 377 AS; T-SOX5: 347 AS). Die für die SOX-Proteinfamilie typische HMG-Box ist in allen Proteinvarianten vorhanden. Unterschiede zwischen den Proteinisoformen bestehen im N-terminalen Bereich, der nicht in T-SOX5 vorhanden ist. Die für die D-Gruppe der SOX-Proteine (SOX5, SOX6) typischen N-terminalen Motive wie Leucin-Zipper Motiv, Coiled-Coil Domäne und Q-Box (Glutamin-reiche Region) sind in der Protein-Protein-Interaktion und Transkriptionsfaktor-Funktion beteiligt (Ikeda et al., 2002).

Für Mensch und Maus konnte per Northern-Blot bereits gezeigt werden, daß das kurze Transkript nur in Proben aus Hodengewebe zu detektieren ist. Für die Maus konnte darüber hinaus gezeigt werden, daß Sox5-Antikörper nur die Zellkerne adulter, postmeiotischer Keimzellen, also von Spermatiden, anfärben (Denny et al., 1992; Connor et al., 1994; Wunderle et al., 1996; Ikeda et al., 2002). Durch die Vervierfachung des *MSY2* Repeats in der Human *DDX3Y* Promoterregion ist auch die SOX5-TFBS vierfach stromaufwärts zu *DDX3Y* T-TSS-I vorhanden.

---

[38] 1. GenBank Acc.No. NM_006940, 763AS, 84 kDa
2. GenBank Acc.No. NM_152989, 750 AS, 82,4 kDa
3. GenBank Acc.No. NM_178010, 377 AS, 42 kDa

[39] GenBank Acc.No. S83308, 347 AS, 38,8 kDa

Alternativ zu SOX5 wurden durch die Programme MatInspector und TRANSFAC die SOX-TFs SOX30 (H-Gruppe SOX-TFs) und SOX9 (E-Gruppe SOX-TFs) vorhergesagt. SOX30 wird als Gonaden-spezifischer SOX-TF mit Keimzelldifferenzierungsfunktion beschrieben und ist nur für die *DDX3X* Promoterregion vorhergesagt (Osaki, et al., 1999; Han et al., 2010). SOX30 bindet wahrscheinlich mit höherer Affinität, als SOX5, an den *DDX3X*-Promoterabschnitt da Matrix-Wert (0,987) und Optimized-Wert (0,92) höher sind. SOX9 bindet wahrscheinlich mit etwas geringerer Affinität (Matrix-Wert = 0,966) an die genomischen *DDX3Y* und *DDX3X* Promoterregionen. SOX9 wird in der männlichen Keimbahn nur in Sertoli-Zellen exprimiert (HPA; Sinclair, 2001; Vidal et al., 2001; Barrionuevo et al, 2006; Kocer et al., 2009; Sekido, 2010; Veitia, 2010). Da die meisten SOX-TFs das DNA-Motiv „ACAAT“ erkennen (Harley et al., 1994; Mertin et al., 1999; Sekido, 2010), ist eine Bindung unterschiedlicher SOX-Faktoren zu unterschiedlichen Entwicklungszeitpunkten oder in unterschiedlichen Geweben möglich.

DMRT

Als zweite Faktorfamilie ist die DMRT („Doublesex- and mab-3-related transcription factor“) Familie für diese Position in *MSY2*/*MSY2-X* vorhergesagt. Von DMRT Faktoren ist bekannt, daß sie in vielen verschieden Vertebratenspezies entscheidend bei der Keimbahnentwicklung und Spermatogenese beteiligt sind (Õunap et al., 2004; Hong et al., 2007; Kawamata et al., 2007; Kim et al., 2007; Matsuda et al., 2007; Murphy et al., 2007; Waters et al., 2007; Graves, 2008; Veyrunes et al., 2008; Wallis et al., 2008). Die vorhergesagte Bindungsaffinität, der Matrix-Wert, ist allerdings geringer als für SOX5. Die Genomatix „Matrix Library“ (Vers. 8.2) gibt für die Faktoren der DMRT TF-Familie an, daß diese präferentiell mit Keimzellen, embryonalem Gewebe, sowie Hoden- und Ovargewebe assoziiert sind.

RXR/VDR

Als weitere konservierte TFBS ist das „Vitamin D Response-Element“ zu nennen. Das aktive Vitamin D_3 bindet an den „Vitamin D Rezeptor“ (VDR), der mit dem „Retinoid X Rezeptor“ (RXR) heterodimerisiert, und schließlich an das Vitamin D Response-Element bindet (Kato, 2000; Norman, 2006). Für Mensch und Ratte konnte eine Bedeutung des VDRs für die normale Keimbahnentwicklung und Spermatogenese gezeigt werden (Johnson et al., 1996; Kinuta et al., 2000; Corbett et al., 2006; Aquila et al., 2008; Blomberg Jensen et al., 2010). Die TFBS liegt sowohl im *DDX3Y* als auch im *DDX3X* Promoter stromabwärts der hodenspezifischen TSS-Positionen.

YY1

In KRiH-6 ist die cTFBS für YY1 lokalisiert. Das Zink-Finger Protein YY1 kann je nach Promoterkontext verschiedene Aufgaben erfüllen. Bekannt sind Funktionen als Transkriptions-

Aktivator und -Repressor, als „Initiator-Motif Binding-Protein“ und als „Polycomb-Group Protein“ (PcG; Atchison et al., 2003; Liu & Shi, 2005; Gordon et al., 2006; Wilkinson et al., 2006; Liu et al., 2007; Schwartz & Pirrotta, 2007; Xi et al., 2007).
YY1 ist evolutionär von Drosophila bis Mensch konserviert und in der Keimzell-Entwicklung beteiligt. In der Maus verändert sich die *YY1*-Expression deutlich während der verschiedenen Spermatogenese-Stadien (Wu et al., 2009).
Als Inr-Motiv Bindeprotein kann YY1 direkt an die RNA-Pol II und TF-IIB binden und rekrutiert so die basale Transkriptionsmaschinerie an den Initiator. Xi et al., (2007) konnten zeigen, daß YY1 TFBS in Core-Promoterregionen deutlich überrepräsentiert vorkommen. Die YY1 Bindesequenz im *DDX3X* Promoter (BAC RP13-13A3, Position 29.994) lautet 5´-GCTTCACAAAATGGCCACC-3´ und beinhaltet eine reverse Inr-Konsensussequenz (CCATTTT; unterstrichen in YY1 Motiv). Stromaufwärts lokalisierte Bindestellen für den TF SP1 verstärken die Transkriptionsinitiation über YY1 (Lee et al., 1993; Smale & Kadonaga, 2003; Li et al., 2008). Bei der Transkriptionsinitiation über YY1 und SP1 muß die Transkription nicht zwingend an der Inr +1 Stelle beginnen, sondern kann zu alternativen TSSs führen.

SPI1

Als weiterer Faktor mit Keimbahn-Funktion ist SPI1 (PU.1), aus der ETS („erythroblast transformation specific“) TF-Familie, zu nennen. Die zugehörige cTFBS, mit Matrix-Werten oberhalb von 0,85, konnte zwischen KRiH-6 und -7 lokalisiert werden. Olive et al. (2007) konnten eine SPI1 Expression zu Beginn der Keimzell-Spezifizierung in „primordialen Keimzellen“ (PGCs), in SSCs („spermatogonial stem cells“), in Spermatogonien und in pachytänen Spermatozyten zeigen. Als eine der Funktionen von SPI1 wird die Erhaltung eines undifferenzierten Zellstatus in SSCs und frühen Spermatogonien genannt (Olive et al., 2007). In der *DDX3Y* Promoterregion ist die funktionell verwandte SPI-B TFBS in Exon-Tb lokalisiert. In der *DDX3X* Promoterregion liegt die TFBS vor der T/B-TSS-II.

AP1

Der in CNSiH-1 lokalisierte TF AP-1, auch JUN genannt, scheint in verschiedenen Signalwegen im Hoden involviert zu sein. Es werden eine Rolle bei der Apoptose, Funktionen während der Spermatogenese, Verbindungen zum Retinolsäure-Signalweg im Hodengewebe, sowie Promoteraktivierungsfunktionen in Sertoli-Zellen beschrieben (Cohen et al., 1993; Benkoussa et al., 2002; Suomalainen et al., 2004; Liu et al., 2006; Vernet et al., 2006; Kaur et al., 2008; Wang et al., 2010).

MESP1 & SCX

Die in CNSiH-1 lokalisierten TFs MESP1 („Mesoderm posterior protein 1“) und SCX („SCLERAXIS“) gehören zur Familie der bHLH-TFs. In Maus Hodengewebe konnte MesP1 in reifen Spermatiden (Saga et al., 1996) und Scleraxis in Sertoli-Zellen identifiziert werden (Muir et al., 2008).

MYC/MAX

Der bHLH-Heterodimerkomplex der TFs MYC („Proto-oncogene c-Myc“) und MAX („Myc-associated factor X“) gilt als transkriptioneller Aktivator der an eine E-Box (5´-CANNTG-3´) binden kann und in Zellzyklusprogression, Differenzierung und Apoptose involviert ist (Blackwood & Eisenman, 1992; Amati & Land, 1994). Die Proteinexpression beider TFs in Hodengewebe ist in HPA zu sehen.

Die Analyse der cTFBS zeigt, daß für beide Promoterregionen TFs mit entscheidenden Funktionen für die Keimbahn-Entwicklung, Spermatogenese und Chromatin-Looping bzw. Chromatin-Remodelling binden können. Auf Grund der Konserviertheit und der hohen, *in-silico* ermittelten, Matrix-Werte der potentiellen TFBS ist eine Funktionalität der vorgestellten TFs möglich.

4.6.3.2. TFBS-Mustererkennung in den *DDX3Y* und *DDX3X* Promoterregionen

Im zweiten Abschnitt der TF-Analyse werden TFBS mit deutlicher Über- oder Unterrepräsentation in den Human *DDX3Y* und *DDX3X* Promoterregionen identifiziert. Mit dieser Analyse sollen erstens Unterschiede in der Häufigkeit der kartierten Transkriptionsfaktoren zwischen den *DDX3Y* und *DDX3X* Promoterregionen erkannt werden und zweitens sollen TFs erkannt werden, die im Vergleich zu einem von Genomatix standardisierten Promoter über- oder unterrepräsentiert sind. Als Background-DNA wird von dem Programm eine gleichlange Nicht-Promotersequenz desselben Organismus gewählt. Als Vergleichspromoter nutzt Genomatix alle verfügbaren Genomatix-Promotersequenzen ab ~ 600bp Länge (500bp stromaufwärts und 100bp stromabwärts der TSS).

MatInspector gruppiert für die Auswertung identische oder funktionell verwandte TFs zu Matrix-Familien zusammen. Die Matrix-Familie der SRY/SOX-Faktoren umfaßt z.B. für die *DDX3Y* Promoterregion die TFs SRY, SOX5, SOX9, HMGA („High Mobility Group Protein A1/2“ = HMGI(Y)/(C); Chromatin-Architektur und Chromatin-Remodelling) und HBP1 („HMG box-containing protein 1“; Transkriptions-Repressor) die alle eine Expression in Hodengewebe aufweisen. In der *DDX3X* Promoterregion besteht die SRY/SOX-TF-Familie aus den TFs SRY,

SOX4, SOX5, SOX9, SOX30 und HMGA. Die HMGA-Proteine sind vor allem bei Prozessen der zellulären Differenzierung in undifferenzierten Zellen anzutreffen (Bussemakers et al., 1991; Chiappetta et al., 1996; Reeves, 2000). Die Anzahl der TFBS einer TF-Familie variiert ebenfalls stark zwischen den beiden Promoterregionen. So sind in der *DDX3Y* Promoterregion 21 TFBS aus der SRY/SOX-TF-Familie kartiert. Für die *DDX3X* Promoterregion sind es 12 SRY/SOX-TF-Faktoren.
In Tabelle 4-11 und 4-12 sind alle TF-Familien erfaßt, die einen statistisch signifikanten Z-Score oberhalb von 2 oder unterhalb von -2 aufweisen. Ein signifikanter Z-Score entspricht einem p-Wert von 0,05. Der Z-Score errechnet sich aus der Formel: $\boldsymbol{Z} = (\boldsymbol{X} - \boldsymbol{E} - \boldsymbol{0.5})/\boldsymbol{S}$. Dabei bezeichnet X die Anzahl gefundener TFBS in einer gegebenen Sequenz, E ist der Erwartungswert und S ist die Standardabweichung (Ho Sui, et al., 2005). Die Z-Scores werden mit einer Stetigkeitskorrektur berechnet.

Die Auswertung ergab für die *DDX3Y* Promoterregion 22 TF-Familien (18 über- ↑, 4 unterrepräsentiert ↓), die von einer normalen Verteilung in genomischen bzw. Promoterregionen abweichen (Tab. 4-11). Für 13 (9↑, 4 ↓) der 22 TF-Familien ist eine Keimbahn-Expression bekannt. Für die TF-Familien NRF1 (↑; „Nuclear Respiratory Factor 1“), HES (↑; Homologe TFs zu Drosophila „hairy and enhancer of split“), ZBPF (↓; „Zinc binding protein factors“), SP1 (↓; „Stimulating Protein 1“) und EGRF (↓; „Early Growth Response Factors“) ist eine Promoter-Assoziation bekannt, d.h. daß diese Faktoren in Promoterregionen zweimal häufiger vorkommen als in anderen genomischen Regionen. In Vertebraten sind dies häufig (G+C)-reiche TFBSs, auf Grund des höheren (G+C)-Gehaltes in den Promoterregionen. Die Promoter-Assoziation ist unabhängig von der eingegebenen Sequenz und ist eine Eigenschaft der jeweiligen TF-Familie. Die fünf TF-Familien DMRT, DLXF („Distal-less homeobox transcription factors“), CREB, BRAC („Brachyury gene“) und SRY/SOX weisen die höchsten Z-Scores (Promoter) in der *DDX3Y* Promoterregion auf.

Für die *DDX3X* Promoterregion konnten 38 (26↑, 12 ↓) TF-Familien mit auffälliger Verteilung ermittelt werden. Davon ist für 21 (12↑, 9 ↓) TF-Familien eine Keimbahn-Expression bekannt (Tab. 4-12). Die TF-Familien NRF1 (↑), ZF5 (↑, „Zinc finger/POZ domain transcription factor“), MTE (↑, Core-Promoter „Motif Ten Element“), NUDR (↑, „nuclear DEAF-1 related transcriptional regulator protein“), EGRF (↓), ZBPF (↑), CDE (↑, „Cell cycle dependent element“), MAZF (↑, „Myc associated Zinc fingers“), CTCF (↑, „CCCTC-binding factor“, „CTCF and BORIS insulator protein“), HES (↑), PLAG (↑, „Pleomorphic adenoma gene“), OAZF (↑, „Olfactory associated zinc finger protein“), AP2 (↑, „Activator protein 2“), SP1 (↑), XCPE (↑, Core-Promoter „X gene core promoter element 1“) und PAX9 (↓, „Paired box protein 9“) zeigen dabei eine Promoter-Assoziation.

Tab. 4-11: Über- und Unterrepräsentierte TFBS in der *DDX3Y* Promoterregion.

In Spalte 1 werden die TF-Familien angegeben. Spalte 2 gibt die gefundenen TFBS einer TF-Familie an. In Spalte 3 wird die erwartete TFBS-Anzahl einer gleichlangen Background DNA-Sequenz angegeben. Spalte 5 und 9 geben die Überrepräsentation einer TFBS an (Genom oder Promoter): x-fache Anzahl gefundener TFBS einer Familie zu erwartetem Wert. Spalte 6 gibt den Z-Score für eine Background-DNA an. Die Spalten 7, 8, 9 und 10 sind, entsprechend den Spalten 3,4, 5 und 6, für die eingegebene *DDX3Y* Promotersequenz. In Spalte 11 wird die Promoter-Assoziation der TF-Familien angegeben. Grün markierte Zeilen markieren TF-Familien mit bekannter Expression in Hoden-gewebe. Die Tabelle ist verkürzt auf Faktoren, die einen signifikanten Z-Score oberhalb von 2 bzw. unterhalb von -2 aufweisen. Die Tabelle ist nach absteigendem Z-Score (Spalte 5, Genom) sortiert.

1	2	3	4	5	6	7	8	9	10	11
TF Familie	**Anzahl TFBS (=X)**	**Erwartet (Genom) (=E)**	**Std. Abw. (=S)**	**Überrepräsentation (Genom)**	**Z-Score (Genom)**	**Erwartet (Promoter) (=E)**	**Std. Abw. (=S)**	**Überrepräsentation (Promoter)**	**Z-Score (Promoter)**	**Promoter-Assoziation**
E2F	14	3,37	1,83	4,15	5,52	6,58	2,56	2,13	2,7	nein
CREB	21	8,72	2,95	2,41	4	8,66	2,94	2,43	4,03	nein
DMRT	19	8,3	2,88	2,29	3,55	5,04	2,24	3,77	6	nein
HIF	5	1,18	1,09	4,23	3,05	2,24	1,5	2,23	1,51	nein
BRAC	7	2,09	1,44	3,35	3,05	1,53	1,24	4,59	4,03	nein
FXRE	4	0,86	0,93	4,66	2,85	0,87	0,93	4,61	2,83	nein
PAX3	4	0,96	0,98	4,16	2,59	1,37	1,17	2,93	1,83	nein
NRF1	2	0,25	0,5	7,98	2,49	3,97	1,99	0,5	-1,24	ja
NKX1	8	3,48	1,86	2,3	2,16	1,99	1,41	4,02	3,91	nein
GRHL	5	1,74	1,32	2,88	2,1	1,82	1,35	2,75	1,99	nein
HES	5	1,76	1,33	2,84	2,06	3,56	1,88	1,41	0,5	ja
DLX	13	7,36	2,71	1,77	1,9	4,16	2,04	3,12	4,09	nein
SRY/SOX	21	15,33	3,9	1,5	1,84	9,92	3,14	2,32	4	nein
PARF	17	10,99	3,31	1,55	1,67	7,01	2,64	2,42	3,59	nein
HOMF	31	26,23	5,09	1,18	0,84	17,44	4,16	1,78	3,14	nein
TBP	17	14,54	3,8	1,17	0,52	8,6	2,93	1,98	2,7	nein
FKHD	18	16,76	4,08	1,07	0,18	10,17	3,18	1,77	2,3	nein
ZBPF	4	3,36	1,83	1,19	0,08	11,68	3,41	0,34	-2,4	ja
SP1	3	2,45	1,57	1,22	0,03	8,89	2,98	0,34	-2,15	ja
LHXF	20	21,03	4,57	0,95	-0,34	11,72	3,42	1,71	2,28	nein
EGRF	2	2,73	1,65	0,73	-0,74	10,37	3,21	0,19	-2,76	ja
OCT1	13	23,03	4,78	0,56	-2,2	13,49	3,66	0,96	-0,27	nein

Tab. 4-12: Über- und Unterrepräsentierte TFBS in der *DDX3X* Promoterregion.

1	2	3	4	5	6	7	8	9	10	11
TF Familie	Anzahl TFBS (=X)	Erwartet (Genom) (=E)	Std. Abw. (=S)	Überrepräsentation (Genom)	Z-Score (Genom)	Erwartet (Promoter) (=E)	Std. Abw. (=S)	Überrepräsentation (Promoter)	Z-Score (Promoter)	Promoter-Assoziation
NRF1	15	0,25	0,5	59,82	28,46	3,97	1,99	3,78	5,28	ja
ZF5	15	0,34	0,58	43,94	24,23	2,71	1,65	5,54	7,17	ja
MTE	7	0,45	0,67	15,59	9,03	2,71	1,65	2,58	2,3	ja
NUDR	5	0,38	0,62	13,13	6,67	1,27	1,13	3,93	2,86	ja
EGRF	13	2,73	1,65	4,76	5,92	10,37	3,21	1,25	0,66	ja
ZBPF	14	3,36	1,83	4,17	5,53	11,68	3,41	1,2	0,53	ja
E2F	14	3,37	1,83	4,15	5,52	6,58	2,56	2,13	2,7	nein
CDE	3	0,19	0,43	16,04	5,35	1,27	1,13	2,37	1,09	ja
MZF1	11	2,49	1,58	4,41	5,08	4,34	2,08	2,54	2,96	nein
MAZF	8	1,49	1,22	5,39	4,94	5,34	2,31	1,5	0,94	ja
CTCF	8	1,57	1,25	5,08	4,72	6,92	2,63	1,16	0,22	ja
HES	8	1,76	1,33	4,54	4,32	3,56	1,88	2,25	2,09	ja
PLAG	5	0,77	0,88	6,46	4,24	2,81	1,68	1,78	1,01	ja
HIF	6	1,18	1,09	5,08	3,97	2,24	1,5	2,68	2,18	nein
OAZF	4	0,65	0,8	6,17	3,54	1,34	1,16	2,99	1,87	ja
AP2	4	0,68	0,83	5,84	3,4	1,94	1,39	2,06	1,12	ja
SP1	8	2,45	1,57	3,26	3,22	8,89	2,98	0,9	-0,47	ja
XCPE	3	0,46	0,68	6,54	3,01	2,66	1,63	1,13	-0,1	ja
XBBF	7	2,37	1,54	2,96	2,69	3,29	1,81	2,13	1,77	nein
SPZ1	4	0,98	0,99	4,06	2,54	1,74	1,32	2,3	1,34	nein
PAX9	2	0,27	0,52	7,49	2,39	0,81	0,9	2,46	0,76	ja
ZF35	3	0,62	0,79	4,84	2,39	1,05	1,03	2,85	1,41	nein
KLFS	13	6,48	2,54	2,01	2,37	11,69	3,41	1,11	0,24	nein
ETSF	16	8,87	2,97	1,8	2,23	10,99	3,31	1,46	1,36	nein
GCMF	4	1,15	1,07	3,49	2,2	1,8	1,34	2,22	1,26	nein
NOLF	4	1,18	1,08	3,4	2,14	2,33	1,53	1,72	0,77	nein
FKHD	9	16,76	4,08	0,54	-2,02	10,17	3,18	0,88	-0,53	nein
HOMF	16	26,23	5,09	0,61	-2,11	17,44	4,16	0,92	-0,47	nein
OCT1	13	23,03	4,78	0,56	-2,2	13,49	3,66	0,96	-0,27	nein
SRY/SOX	12	15,33	3,9	0,46	-2,26	9,92	3,14	0,71	-1,09	nein
GATA	3	9,7	3,11	0,31	-2,32	6,34	2,52	0,47	-1,53	nein
HBOX	8	17,03	4,11	0,47	-2,32	9,97	3,15	0,8	-0,78	nein
MYT1	2	8,4	2,89	0,24	-2,39	5,9	2,43	0,34	-1,81	nein
HOXC	1	6,69	2,58	0,15	-2,4	4,09	2,02	0,24	-1,78	nein
CDX	1	7,15	2,67	0,14	-2,49	4,03	2,01	0,25	-1,76	nein
NKXH	3	12,41	3,51	0,24	-2,82	8,12	2,85	0,37	-1,98	nein
LHXF	8	21,03	4,57	0,38	-2,96	11,72	3,42	0,68	-1,24	nein
HOX	8	21,53	4,62	0,37	-3,04	12,82	3,57	0,62	-1,49	nein

Für die fünf TF-Familien ZF5, NRF1, MZF1 („Myeloid zinc finger 1 factors“), NUDR und E2F konnten in der *DDX3X* Promoterregion die höchsten Z-Scores (Promoter) festgestellt werden.

12 der 22 auffällig verteilten TF-Familien aus der *DDX3Y* Promoterregion sind in der *DDX3X* Promoterregion wiederzufinden (Tab. 4-13). Die TF-Familien E2F, HIF („Hypoxia induced factor“), NRF1 und HES sind in beiden Promoterregionen überrepräsentiert. Die OCT1 („Octamer binding protein 1“) TF-Familie ist in beiden Promoterregionen unterrepräsentiert. Die TF-Familien SRY/SOX, HOMF („Homeodomain transcription factors“), FKHD („Fork head domain factors“), ZBPF, SP1, LHXF („Lim homeobox factors“), und EGRF weisen eine gegenläufige Verteilung auf. Davon haben wiederum 11 eine bekannte Keimbahnexpression. Bis auf die CREB- und DMRT-TF-Familien sind somit alle Keimbahn TF-Familien aus *DDX3Y* auch mit einer auffälligen Verteilung in der *DDX3X* Promoterregion vorhanden.
D.h. also, daß ein bestimmtes Grundset an Faktoren immer in beiden Promoterregion gehäuft aufzutreten scheint. Für die *DDX3X* Promoterregion fällt allerdings auf, daß wesentlich mehr TF-Familien ein erhöhtes oder geringeres Vorkommen als erwartet aufweisen und weit mehr TF-Familien eine Promoter-Assoziation aufweisen, als dies für die *DDX3Y* Promoterregion der Fall ist.

Tab. 4-13: Vergleich der über- und unterrepräsentierten TF-Familien in den *DDX3Y* und *DDX3X* Promoterregionen.
--- = keine signifikante Veränderung. TF-Familien mit Expression in Hodengewebe sind grün markiert.

TF-Familie	*DDX3Y*		*DDX3X*		Promoter-assoziation
	Genom-Seq.	Promoter-Seq.	Genom-Seq.	Promoter-Seq.	
E2F	↑	↑	↑	↑	---
HIF	↑	---	↑	↑	---
NRF1	↑	---	↑	↑	ja
HES	↑	---	↑	↑	ja
SRY/SOX	---	↑	↓	---	---
HOM	---	↓	↓	---	---
FKHD	---	↓	↓	---	---
ZBPF	---	↓	↑	---	ja
SP1	---	↓	↑	---	ja
LHXF	---	↑	↓	---	---
EGRF	---	↓	↑	---	ja
OCT1	↓	---	↓	---	---

Zusätzlich zu den über- und unterrepräsentierten TF-Familien konnten in der *DDX3X* Promoterregion auffällige lokale Clusterungen für die drei TF-Familien, ZF5, SP1 und MZF1, identifiziert werden.

ZF5

Für ZF5 (auch ZFP161 genannt) konnten von BAC-Position 29.552 bis 29.812 (261 bp; BAC RP13-13A3) 12 potentielle, teilweise überlappende Bindestellen kartiert werden. Diese ZF5 Region beginnt direkt vor der *MSY2-X* Sequenz (BAC-Position 29.570 - 29.676) und endet vor der *DDX3X* T-TSS-III (BAC-Pos. 29.825) im S-Trakt.

ZF5 gehört zu der BTB/POZ TF-Familie (broad complex, tramtrack, bric-à-brac/ poxvirus and zinc finger; Sugiura et al., 1997) und verfügt über eine BTB/POZ-Domäne (N-term) und fünf C2H2-Zinkfinger-Domänen (C-term). Proteine mit BTB/POZ-Domäne und gleichzeitig C2H2 Zinkfingern werden auch POK-TFs (von POZ und Krüppel-Typ Zink-Finger) genannt. Die Konsensus DNA-Bindesequenz 5'-GC/GGCGCGA/G-3' wurde von Obata et al. (1999) festgestellt. Orlov et al. (2007) konnten zeigen, daß ZF5 auch generell $(GCC)_n$-Tripletts, wie z.B. in der 5´-Region des FMR1-Gens, erkennt und an diese bindet.

ZF5 geht sowohl als Homomer, als auch als Heteromer Proteininteraktionen ein (Numoto et al., 1999). Protein-Protein Interaktion wird von beiden Domänen, POZ und Zink-Finger, vermittelt, was daher zu den unterschiedlichen Wirkungen als Repressor oder Aktivator in Kombination mit verschieden Co-Faktoren führt (Bardwell & Treisman, 1994; Kaplan & Calame, 1997; Ahmad et al., 1998; Collins et al., 2001; Lee & Lee, 2004; Wang et al., 2005).

SP1

Eine zweite Region mit repetitiven TFs betrifft den Faktor SP1 („Stimulating Protein 1“). Von Position 29.681 bis 29.932 (252 bp) konnten fünf SP1 TFBS kartiert werden. Diese SP1-Region erstreckt sich direkt stromabwärts von *MSY2-X* bis in den Y-Trakt. Wie oben bereits beschrieben, verstärken sich GC-Boxen und TC-Boxen (Kapitel 4.6.2), wie sie im Y-Trakt zu finden sind, gegenseitig. Wie bei der ZF5-Region ist die Lage mit der (C,G)-reichen Region, zwischen *MSY2-X* und dem 5´-Transkriptende, korreliert.

Das SP1- Bindemotiv 5´-GGGCGG-3´ wird als GC-Box bezeichnet. In CpG-Promotoren erfolgt die Rekrutierung des Präinitiationskomplexes maßgeblich durch Aktivator-TFs wie SP1 und nicht durch bestimmte Core-Promoter-Motive. In TATA-losen Promotoren liegen die TSSs meist 36 bis 80 bp abwärts multipler GC-Boxen, was in CpG-Promotoren häufig zu mehreren schwachen TSS führt. Die TSSs können dabei über einen Bereich von 100 und mehr Basenpaaren verteilt sein. Eine SP1 induzierte Transkriptionsinitiation ist häufig ein Anzeichen für ein konstitutiv exprimiertes „Housekeeping Gen“ oder für ein gewebespezifisch exprimiertes Gen. Durch Methylierung des ersten und zweiten Cytosins auf dem Gegenstrang 5´-CCGCCC-3´ ($^{m}Cp^{m}CpG$) kann die Genexpression SP1-regulierter Gene gewebespezifisch durch ein epigenetisches Ereignis reguliert werden. Durch eine solche Methylierung wird die Bindungsfähigkeit von SP1 an die Promotersequenz unterbunden und es kann keine Transaktivierung stattfinden (Clark et al., 1997; Suske, 1999; Jiang et al., 2006; Douet et al.,

2007). Die Methylierung einer GC-Box kann zudem Ausgangspunkt einer größeren Methylierungsausbreitung in einer Promoterregion sein.

Es konnte gezeigt werden, daß hodenspezifisch alternativ gesplicte SP1-Transkripte und verschiedene SP1-Proteinvarianten existieren. Diese weisen ein Zelltyp- und Spermatogenese-stadium-spezifisches Expressionsprofil auf (Thomas et al., 2007; Widlak et al., 2007). Saffer et al. (1991) haben an Mausstudien gezeigt, daß die höchsten SP1-Transkript Expressionslevel in Spermatiden zu finden sind.
SP1 kann direkt mit TAFs aus dem PIC interagieren, die wiederum mit HATs („histone acetyl transferases“) und TBP interagieren. Selektive SP1-Promoteraktivierung kann durch Zelltyp- und Spermatogenesestadium-spezifische Co-Aktivatoren/Co-Faktoren, wie die TAFs, erfolgen, was zu einem spezifischen Gen-Expressionsmuster während der Spermatogenese führt (Thomas et al., 2007). Die verwandte GT-Box (5´-GGTGTGGGG-3´) und GC-Boxen können auch einen Repressoreffekt auf die Promoteraktivität haben, wenn repressive SP-TFs wie SP3 an diese binden (Kingsley & Winoto, 1992; Hamann et al., 1994; Young & Mendelson, 1997; Zhao & Meng, 2005).

MZF1

Eine dritte Region wird durch die acht TFBSs des TFs „Myeolid Zinc Finger 1“ (MZF1) konstituiert. Diese ZnF-Region ist in BAC-Position 29.813 bis 30.090 der *DDX3X* Promoter-region zu finden. Die Region beginnt direkt vor T-TSS-III und endet ~ 110 Nukleotide stromabwärts des Y-Traktes. Auch hier konzentrieren sich die TFBS in den (C,G)- bzw. (C,T)-reichen Regionen um den Hoden-spezifischen *DDX3X* Transkriptstart. D.h., daß die identifizierten Sequenzblöcke in der *DDX3X* Promoterregion nicht nur Einfluß auf die Promoter-Architektur haben können, sondern auch massiv die Bindung einiger TFs beeinflussen, die lokal und repetitiv vorkommen.
MZF1 ist nur schwach in Keimzellen exprimiert, wird aber stark in Leydig-Zellen synthetisiert (HPA). Von MZF1 ist bekannt, daß es eine wichtige Rolle bei Zellproliferation und Tumorentwicklung einnimmt (Gaboli et al., 2001).

Eine vergleichbare lokale Akkumulation von Faktoren konnte in der potentiellen *DDX3Y* Promoterregion nicht gefunden werden. Durch die Vervierfachung der *MSY2*-Sequenzeinheit kommen die dort kartierten Faktoren, wie SOX5 und DMRT, allerdings vier Mal mit einem Intervall von ~ 100 bp vor.

4.6.3.3. TF-Module in den Human *DDX3Y* und *DDX3X* Promoterregionen

In diesem Unterkapitel werden potentielle TF-Module in den *DDX3Y* und *DDX3X* Promoterregionen vergleichend betrachtet. Für die Erkennung von TF-Modulen wurde das Genomatix-Programm „ModelInspector“ verwendet (Vers. 5.6.6; Frech et al., 1997; Klingenhoff et al., 1999). Die zu Grunde liegende TF-Datenbank ist für die Genomatix-Programme stets die „Matrix Family Library“ (Vers. 8.2). Die Einstellungen für die TF-Modul Erkennung wurden so gewählt, daß nur 100% vollständige TF-Module gesucht werden.

In den untersuchten 2,5 kb *DDX3Y* und *DDX3X* Promotersequenzen konnten 39 potentielle *DDX3Y* und 38 potentielle *DDX3X* TF-Module identifiziert werden (Tab. 4-14 & 4-15). Für 18 der *DDX3Y* TF-Module sind die jeweiligen TFs in Hodengewebe exprimiert. Für *DDX3X* sind es 24 TF-Module mit Hodenexpression der einzelnen TFs. Eine Vorhersage, ob sich die potentiellen TF-Komplexe tatsächlich in Keimzellen bilden ist *in-silico* nicht möglich. Sechs der TF-Module (SRY-SPI1, XBP1-USF, SOX-PAX6, SOX-OCT, SPI1-AP1, SP1-SPI1) konnten an unterschiedlicher Position und in unterschiedlicher Anzahl in beiden Promoterregionen identifiziert werden (Abb. 4-49). Hervorzuheben sind die potentiellen SOX-PAX TF-Module, bei denen die X-Y konservierte SOX TF-Bindestelle ein Bestandteil ist. In der *DDX3Y* Promoterregion entstehen zwei SOX-PAX6 TF-Module an den Grenzen von *MSY2-2/3* und *MSY2-3/4*. Diese TF-Module sind durch die Lage in *MSY2-3* und *-4* humanspezifisch. In der *DDX3X* Promoterregion entsteht positionshomolog ebenfalls ein potentielles SOX-PAX TF-Modul an der 3´Grenze der *MSY2-X* Sequenz. Das *DDX3X* SOX-PAX TF-Modul ist, mit Ausnahme von *Macaca*, Primaten-spezifisch und nicht in Rind und Maus zu finden. Zudem fällt auf, daß in beiden Promoterregionen ein TF-Modul mit der konservierten TF-Bindestelle für SPI1 (SPI-B) kartierbar ist. In der *DDX3Y* Promoterregion ist es ein SP1-SPI1 TF-Modul, in *DDX3X* ein SPI1-AP1, bzw. ein SRY-SPI1 TF-Modul.

SOX-PAX sowie SOX-OCT TF-Komplexe sind für viele Entwicklungsprozesse von entscheidender Bedeutung (Reményi et al., 2004; Kondoh & Kamachi, 2010; Wegner, 2010). SOX2-PAX6 TF-Komplexe z.B. sind für die Augenentwicklung von Bedeutung (Kondoh et al., 2004; Inoue et al., 2007). Die Interaktion zwischen SOX- und PAX-Proteinen geschieht zwischen der SOX-HMG- und der PAX-Paired-Domäne (Lang & Epstein, 2003; Lang et al., 2007). Es wird angenommen, daß die DNA dazwischen liegt (persönliche Mitteilung von D. Lang). SOX-TFs binden an die kleine Furche und PAX-TFs an die große Furche der DNA. PAX/SOX Komplexe aktivieren die Transkription stärker, als es den einzelnen Faktoren möglich wäre. Der synergistische Effekt tritt nur auf, wenn beide Bindestellen nah beieinander liegen (~ 12 bp; persönliche Mitteilung von D. Lang).

Tab. 4-14: Auflistung der potentiellen TF-Module in der Human *DDX3Y* Promotersequenz.

In Spalte 1 sind die TFs angegeben, die das pot. TF-Modul bilden. Graue Markierung gibt an, daß TF-Modul auch in der *DDX3X* Promoterregion kartiert wurde. Spalte 2 gibt die Position des TF-Moduls innerhalb der 2.545bp langen (BAC-Pos. 52.646 - 55.190) *DDX3Y* Sequenz an. In Spalte 3 und 4 sind die TF-Core- und TF-Matrix-Werte angegeben. Spalte 5 gibt die Abstände zwischen den TFs an. Spalte 6 gibt eine bekannte Hodenexpression der TFs an. Grün = Expression beider Faktoren im Hodengewebe, gelb = Expression eines Faktors. K. A. = keine Angaben zur Gewebeexpression.

DDX3Y	**Position & Orientierung**	**Core-Wert**	**Matrix-Wert**	**Abstand der TFBS**	**Bekannte Hoden-expression**
IRF1-SPI1	129 - 105 (-)	0,759/1,000	0,750/0,990	4 bp	ja/ja
SPDEF-SPI1	133 - 105 (-)	1,000/1,000	0,896/0,990	8 bp	k. A./ja
SRY-SPI1	157 - 105 (-)	1,000/1,000	0,910/0,990	31 bp	ja/ja
SRY-PAX3	157 - 124 (-)	1,000/1,000	0,910/0,802	13 bp	ja/k. A.
SP1-NF1-TBP	363 - 414 (+)	0,754/1,000/1,000	0,758/0,892/0,933	9 bp/27 bp	ja/ja/ja
TEF-ESRRB	393 - 520 (+)	0,818/1,000	0,758/0,934	110 bp	k. A./ja
BRN2-RXR/VDR	409 - 458 (+)	1,000/1,000	0,901/0,776	28 bp	k. A./ja
XBP1-HNF1	542 - 572 (+)	1,000/1,000	0,917/0,866	12 bp	k. A./k. A.
MYC/MAX-SREBP	557 - 284 (-)	1,000/1,000	0,988/0,739	260 bp	ja/ja
XBP1-USF	562 - 542 (-)	1,000/1,000	0,917/0,985	0 bp	k. A./ja
CEBPB-USF	575 - 546 (-)	0,701/1,000	0,763/0,985	16 bp	ja/ja
NFAT-GATA1	776 - 826 (+)	1,000/1,000	0,926/0,985	35 bp	k. A./ja
HNF1-GATA1	874 - 901 (+)	1,000/1,000	0,919/0,929	13 bp	k. A./ja
HMBOX1-COUP	1.174 - 1.235 (+)	1,000/1,000	0,940/0,802	41 bp	ja/ja
HMBOX1-CEBPA	1.190 - 1.129 (-)	1,000/0,714	0,949/0,797	46 bp	ja/k. A.
TBX5-TBX5	1.356 - 1.574 (+)	1,000/1,000	0,964/0,964	198 bp	ja/ja
SOX5-PAX6	1.395 - 1.435 (+)	1,000/1,000	0,987/0,686	20 bp	ja/ja
SOX5-PAX9	1.417 - 1.349 (-)	1,000/1,000	0,987/0,920	46 bp	ja/k. A.
OCT3-SOX5	1.426 - 1.395 (-)	0,784/1,000	0,878/0,987	12 bp	ja/ja
ATF-HNF1	1.491 - 1.462 (-)	1,000/1,000	0,934/0,798	11 bp	ja/k. A.
SOX5-PAX6	1.494 - 1.534 (+)	1,000/1,000	0,987/0,686	20 bp	ja/ja
OCT3-SOX5	1.525 - 1.494 (-)	0,784/1,000	0,878/0,987	12 bp	ja/ja
HBP1-SPDEF	1.574 - 1.626 (+)	1,000/1,000	0,905/0,958	31 bp	ja/k. A.
ATF-HNF1	1.590 - 1.561 (-)	1,000/1,000	0,934/0,798	11 bp	ja/k. A.
IRF4-SPDEF	1.601 - 1.626 (+)	1,000/1,000	0,832/0,958	5 bp	k. A./k. A.
SPDEF-IRF4	1.626 - 1.601 (-)	1,000/1,000	0,958/0,832	5 bp	k. A./k. A.
SMAD3-AP1	1.677 - 1.875 (+)	1,000/1,000	0,985/0,794	189 bp	ja/ja
ARNT-SPI1	1.739-1.767 (+)	1,000/1,000	0,974/0,926	10 bp	ja/ja
NMYC-MYC/MAX	1.741 - 1.784 (+)	1,000/0,790	0,983/0,874	31 bp	k. A./ja
SPI1-AP1	1.747 - 1.778 (+)	1,000/1,000	0,926/0,812	16 bp	ja/ja
NMYC-SREBP	1.753 - 1.474 (-)	1,000/0,739	0,983/0,645	266 bp	k. A./ja
NF-κB/NKX3-2	1.774 - 1.753 (-)	1,000/1,000	0,705/0,974	6 bp	ja/ja
RUNX2-CEBPB	1.864 - 1.605 (-)	1,000/1,000	0,887/0,899	245 bp	k. A./ja
TBX5-TBX5	1.910 - 1.820 (-)	1,000/1,000	0,993/0,883	70 bp	ja/ja
CEBPB-AP1	1.971 - 1.943 (-)	1,000/1,000	0,924/0,754	16 bp	ja/ja
STAT5-AR	2.019 - 2.057 (+)	1,000/0,868	0,977/0,810	20 bp	k. A./ja
NR4A1/2/3-STAT5	2.048 - 2.021 (-)	1,000/1,000	0,752/0,974	11 bp	k. A./k. A.
KLF4-E4BP4-KLF4	2.087 - 2.033 (-)	1,000/0,747/1,000	0,899/0,769/0,946	20 bp/18 bp	ja/k. A./ja
SP1-SPI1	2.087 - 2.059 (-)	0,872/1,000	0,825/0,953	11 bp	ja/ja

Tab. 4-15: Auflistung der potentiellen TF-Module in der Human *DDX3X* Promotersequenz.

In Spalte 1 sind die TFs angegeben, die das pot. TF-Modul bilden. Graue Markierung gibt an, daß TF-Modul auch in der *DDX3Y* Promoterregion kartiert wurde. Spalte 2 gibt die Position des TF-Moduls innerhalb der 2.545bp langen (BAC-Pos. 28.230 - 30.774) *DDX3X* Sequenz an. In Spalte 3 und 4 sind die TF-Core- und TF-Matrix-Werte angegeben. Spalte 5 gibt die Abstände zwischen den TFs an. Spalte 6 gibt eine bekannte Hodenexpression der TFs an. Grün = Expression beider Faktoren, gelb = Expression eines Faktors.

DDX3X	**Position & Orientierung**	**Core-Wert**	**Matrix-Wert**	**Abstand der TFBS**	**Bekannte Hoden-expression**
P53-BRN3	184 - 231 (+)	1,000/1,000	0,936/0,814	27 bp	ja/k. A.
P53-BRN3	217 - 174 (-)	1,000/1,000	0,909/0,895	23 bp	ja/k. A.
SMARCA3-WT1	414 - 913 (+)	1,000/1,000	0,928/0,930	486 bp	ja/ja
CEBPB-STAT1	435 - 406 (-)	1,000/1,000	0,985/0,782	13 bp	ja/ja
OCT1-SOX5	546 - 519 (-)	1,000/1,000	0,845/0,987	8 bp	ja/ja
ILF1-CEBPB	569 - 588 (+)	1,000/1,000	0,967/0,959	4 bp	ja/ja
SRY-SOX5	630 - 601 (-)	1,000/1,000	0,918/0,987	7 bp	ja/ja
NF-κB/c-REL	1.011 - 1.415 (+)	1,000/1,000	0,853/0,903	392 bp	ja/k. A.
IKZF2-AP2	1.075 - 1.046 (-)	1,000/1,000	0,921/0,908	16 bp	k. A./ja
ETS1-SPI1	1.094 - 1.121 (+)	1,000/1,000	0,953/0,924	7 bp	k. A./ja
EGR1-NFAT	1.324 - 1.353 (+)	1,000/1,000	0,787/0,966	12 bp	ja/k. A.
IKZF2-AP2	1.381 - 1.411 (+)	1,000/0,839	0,913/0,889	17 bp	k. A./ja
c-REL-SP1	1.403 - 1.500 (+)	1,000/0,796	0,903/0,811	84 bp	k. A./ja
SOX5-PAX4	1.425 - 1.460 (+)	1,000/0,862	0,982/0,784	15 bp	ja/ja
EGR1-SP1	1.447 - 1.466 (+)	1,000/1,000	0,908/0,997	4 bp	ja/ja
SP1-SP1	1.452 - 1.472 (+)	1,000/1,000	0,997/0,953	6 bp	ja/ja
EGR1-SP1	1.453 - 1.472 (+)	1,000/1,000	0,993/0,953	4 bp	ja/ja
SP1-SPI1	1.466 - 1.431 (-)	1,000/1,000	0,997/0,898	18 bp	ja/ja
SPI1-SP1	1.589 - 1.619 (+)	1,000/1,000	0,953/0,973	13 bp	ja/ja
SPI1-CEBP	1.609 - 1.437 (-)	1,000/0,857	0,953/0,858	155 bp	ja/k. A.
PEA3-SP1	1.644 - 1.605 (-)	1,000/1,000	0,852/0,973	22 bp	ja/ja
SP1-SPI1	1.678 - 1.724 (+)	0,876/1,000	0,872/0,964	29 bp	ja/ja
SPI1-YY1	1.704 - 1.783 (+)	1,000/1,000	0,964/0,978	59 bp	ja/ja
SREBP-SP1	1.798 - 1.821 (+)	1,000/1,000	0,913/0,894	9 bp	ja/ja
XBP1-USF	1.818 - 1.795 (-)	1,000/1,000	0,940/1,000	0 bp	k. A./ja
STAT-SPI1	1.821 - 1.842 (+)	1,000/1,000	0,944/0,893	2 bp	k. A./ja
IKZF2-AP2	1.837 - 1.805 (-)	1,000/1,000	0,970/0,897	19 bp	k. A./ja
STAT-MITF	1.839 - 1.797 (-)	1,000/0,766	0,944/0,846	26 bp	k. A./k. A.
SP1-SPI1	1.889 - 1.860 (-)	1,000/1,000	0,823/0,954	12 bp	ja/ja
SPI1-AP1	1.913 - 1.934 (+)	1,000/1,000	0,937/0,807	6 bp	ja/ja
SPI1-AP1	1.925 - 1.947 (+)	1,000/1,000	0,960/0,792	7 bp	ja/ja
SRY-SPI1	1.984 - 1.925 (-)	1,000/1,000	0,928/0,960	38 bp	ja/ja
SMAD3-E2F4	2.109 - 2.125 (+)	0,750/1,000	0,715/0,953	3 bp	ja/ja
SMAD3-E2F4	2.124 - 2.108 (-)	0,750/1,000	0,904/0,978	3 bp	ja/ja
PEA3/NF-κB	2.243 - 2.263 (+)	1,000/1,000	0,904/0,978	3 bp	ja/ja
NF-κB/SP2	2.262 - 2.215 (-)	1,000/1,000	0,978/0,849	34 bp	ja/ja
NF-κB/CREB	2.372 - 2.345 (-)	1,000/1,000	0,852/0,971	11 bp	ja/ja
NFAT-BACH1	2.531 - 2.496 (-)	1,000/1,000	0,948/0,982	16 bp	k. A./k. A.

4.6.3.4. Konservierte Abstände zwischen Human cTFBS

Um feststellen zu können in welchen Abschnitten der *DDX3Y* und *DDX3X* Promoterregionen die Abstände zwischen den cTFBS konserviert sind, werden die Abstände sechs ausgewählter cTFBS untereinander und zum Translationsstart ATG verglichen (Abb. 4-49).

Zwischen dem TF SATB1, in KRiH-1, und dem *DDX3Y* ATG liegen 2.112 bp. Dasselbe Intervall in der *DDX3X* Promoterregion hat eine Länge von 1.973 bp. Dies ist ein Unterschied von 140 bp. Unter Berücksichtigung der Vervierfachung des *DDX3Y MSY2* Repeats (308 bp) und der fast vollständigen Reduktion der Pyrimidinsequenz (163 bp) im *DDX3Y* Promoter, ist dieser Längenunterschied von 140 bp zu erklären. Die Abstände zwischen SATB1 und CHR (Y = 372 bp; X = 342 bp), sowie CHR und SOX5 (Y = 528 bp; X = 556 bp) sind dagegen in beiden Promoterregionen stabil. Der Abstand von SOX5 zu YY1 zeigt im Unterschied dazu wieder eine große X-Y Varianz (Y = 581 bp; X = 339 bp). Das Intervall YY1 zu SPI1 zeigt wiederum keinerlei Unterschied (Y = 166 bp; X = 163 bp) und präsentiert sich als sehr stabil. Von SPI1 bis zu HIF1-α sind es im *DDX3Y* Promoter 312 bp, im *DDX3X* Promoter 411 bp. Der Abstand von HIF1- α zum ATG ist wiederum sehr stabil mit 158 bp (Y) und 167 bp (X).

Die Abstandsberechnung der cTFBS zeigt, daß in den Bereichen SOX5 zu YY1 und SPI1 zu HIF1- α die größten Sequenzveränderungen seit Trennung der Geschlechtschromosomen, in Bezug auf die Sequenzlänge, stattgefunden haben. In diesen zwei Sequenzbereichen befinden sich auch auffallender Weise die hodenspezifischen Core-Promoterregionen und Allel-spezifischen Transkriptionsstartstellen beider Gene. Die Abstände zwischen SATB1 zu SOX5 (~900 bp) und zwischen YY1 zu SPI1 (~ 165 bp) sind erhalten geblieben. Somit sind, neben den vielfältigen Allel-spezifischen Veränderungen, auch Bereiche eines ursprünglichen *DDX3*(*X/Y*) Promoters im Humangenom konserviert geblieben.

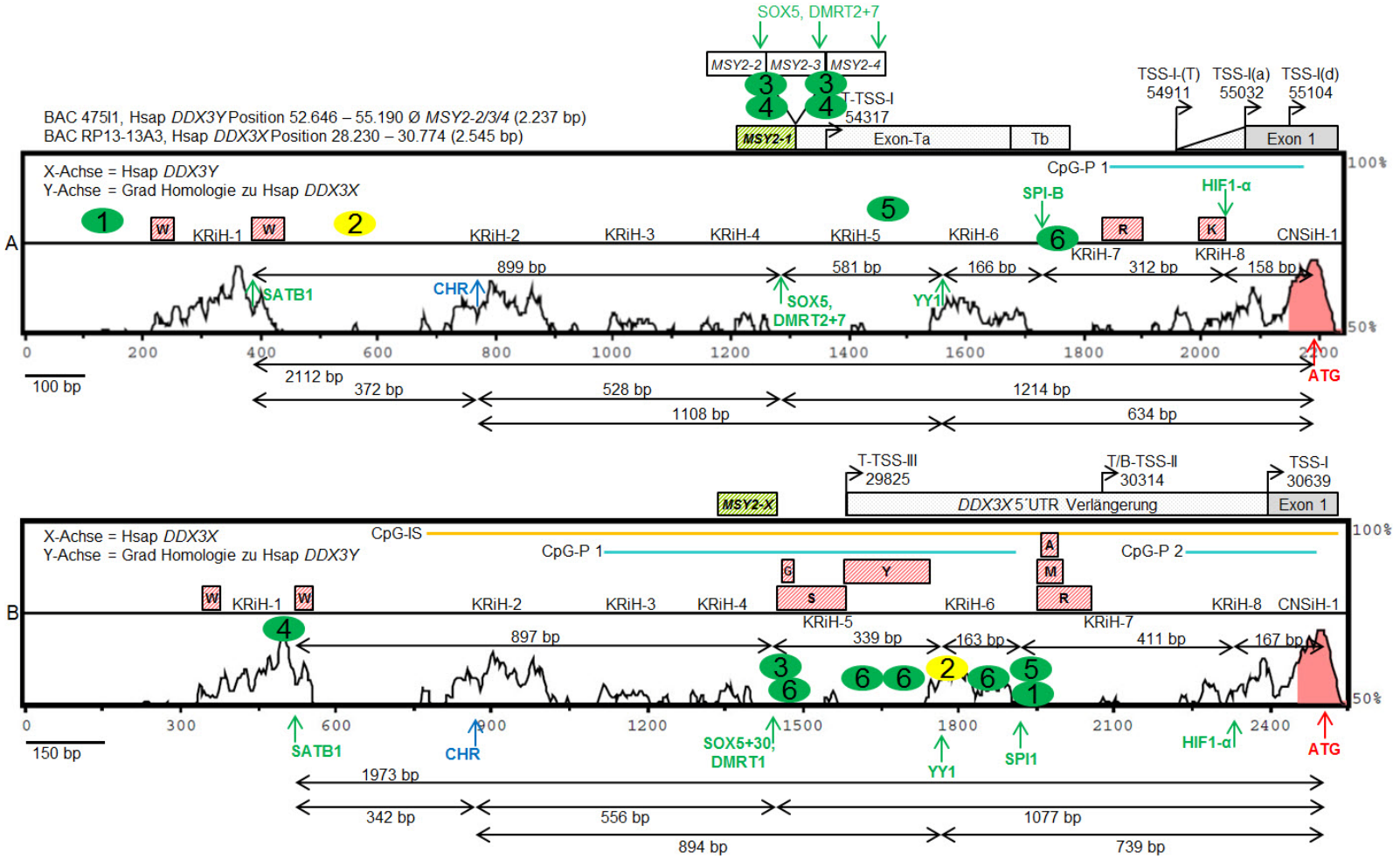


Abb. 4-49: Graphische Darstellung der Abstände zwischen konservierten TFBSs und gemeinsamer TF-Module in der Human *DDX3Y* und *DDX3X* Promoterregion.

A: *DDX3Y* Promoterregion ohne die *MSY2* Repeateinheiten 2, 3 und 4 (308 bp entfernt, Bruch durch Blitz angezeigt und als Einschub oberhalb gezeigt). **B:** *DDX3X* Promoterregion. Die Abstände zwischen den cTFBS sind durch Pfeile markiert. Die Ovale geben TF-Module an, wie in Tab. 4-14 und 4-15 kartiert, die in beiden Promoterregionen vorkommen. Grünes TF-Modul = beide TFs in Hoden bekannt. Gelbes TF-Modul = ein TF in Hoden bekannt. Die TF-Module sind der Reihe nach von 5´ nach 3´ im *DDX3Y* Promoter durchnummeriert. 1 = SRY-SPI1, 2 = XBP1-USF, 3 = SOX-PAX, 4 = SOX-OCT, 5 = SPI1-AP1, 6 = SP1-SPI1.

4.7. Vertikaler Vergleich der Human *DDX3Y* und *DDX3X* Promoterdomänen mit homologen Sequenzen anderer Säugerspezies

Zur Bestätigung der evolutionär konservierten Regionen in den Promoterbereichen der Human *DDX3Y* und *DDX3X* Gene, sowie zur Identifizierung X- und Y-Chromosom-spezifischer Domänen, wurden die homologen X-Y Sequenzen von Schimpanse, Rhesusaffe, *Callithrix*, Rind und Maus ebenfalls mit dem Programm mVISTA auf konservierte Sequenzabschnitte (CNS) analysiert. Dabei erkannte potentielle CpG-Islands und konservierte TF-Bindestellen werden ebenso in den vertikalen Vergleich einbezogen.

Als Ausgangssequenzen für den evolutionären Vergleich werden die bekannten 2,5 kb langen Human *DDX3Y* (2545 bp, BAC-Klon 475I1, Pos. 52.646 - 55.190) und *DDX3X* (2.545 bp, BAC-Klon RP13-13A3, Pos. 28.230 – 30.774) Promotersequenzen verwendet.

***DDX3Y*:**

Folgende genomischen *DDX3Y*/*Ddx3y* Sequenzen wurden für die mVISTA-Analyse verwendet: Schimpanse BAC-Klon „CH251-128L22“[40], *Macaca* BAC-Klon „CH250-541A11“[41], *Callithrix* BAC-Klone „CH259-161K20“[42], Rind BAC-Klon „CH240-56A13“[43], Maus BAC-Klon „RP24-208N6“[44]. Der *DDX3Y/Ddx3y* Sequenzvergleich mit mVISTA mußte für die Primaten, sowie Rind und Maus getrennt durchgeführt werden. Auf Grund der unterschiedlichen Anzahl an *MSY2* Repeats und teils erheblicher Sequenzveränderungen war ein gemeinsamer *DDX3Y*/*Ddx3y* Plot aller sechs Spezies nicht möglich. Für die Primaten konnte nach Entfernung der Human *MSY2* Repeats 3 und 4 ein einheitlicher Plot erstellt werden. Für Rind und Maus wurde je ein einzelner Plot nach Entfernung der Human *MSY2* Repeats 2, 3 und 4 erstellt. Ein einheitlicher Plot aller Spezies führte, vor allem für die Maus *Ddx3y* Promoterregion, zu deutlich abweichenden Ergebnissen. Besonders für die Maus fällt der schnellere Generationswechsel und die damit verbundene höhere Generationszahl seit der Radiation der Säugetiere ins Gewicht, was in der Maus *Ddx3y* Promoterregion zu sehr starken Veränderungen der Promoterstruktur führte.

Primaten:

Im mVISTA-Vergleich der *DDX3Y* Promoterregionen konnten für Schimpanse und *Macaca* durchgehende CNS mit Sequenzhomologien von 98% bzw. 91% zum Human *DDX3Y* Promoter

40 *DDX3Y* Ptro: BAC-Position 61.314 - 58.980; 2.335 bp; revers-komplementär.
41 *DDX3Y* Mmul: BAC-Position 85.273 - 82.948; 2.326 bp; revers-komplementär.
42 *DDX3Y* Cjac: BAC-Position 53.768 - 51.606; 2.163 bp; revers-komplementär.
43 *DDX3Y* Btau: BAC-Position 142.291 - 139.747; 2.545 bp; revers-komplementär.
44 *DDX3Y* Mmus: BAC-Position 96.244 - 93.700; 2.545 bp; revers-komplementär.

festgestellt werden (Tab. 4-16). Für *Callithrix* konnten drei konservierte Sequenzblöcke festgestellt werden, die von einer 5 bp (60 bp in Human) und einer 20 bp (75 bp in Human) langen Lücke unterbrochen werden. Die Sequenzhomologien liegen zwischen 74,5% und 81% (Tab. 4-16). Der graphische Vergleich der Sequenzhomologien ist in Abb. 4-50 einzusehen.

Tab. 4-16: Liste der mit mVISTA ermittelten konservierten Sequenzabschnitte zwischen der Human *DDX3Y* und den Primaten *DDX3Y* Promoterregionen von Schimpanse, *Macaca* und *Callithrix*.

Spalte 2: Position der ermittelten Elemente im Human Y-Chr. BAC 475I1, bezogen auf die verkürzte Sequenz (ohne *MSY2-3/4* = 2.347 bp). Spalte 3: Position der ermittelten Elemente in der 2.347 bp langen Human *DDX3Y* Sequenz (ohne *MSY2-3/4*). Spalte 4: Länge der ermittelten Elemente in der Human *DDX3Y* Promoterregion. Spalte 5: BAC-Positionen der ermittelten Elemente in den Schimpanse, *Macaca* und *Callithrix DDX3Y* Promoterregionen. Spalte 6: Position der ermittelten Elemente in den Primaten *DDX3Y* Sequenzen. Spalte 7: Länge der ermittelten Elemente in den Primaten *DDX3Y* Promoterregionen. Spalte 8: Sequenzidentität der Elemente in %. Werte in Klammern geben die Homologiewerte jeweils für die Schimpansen-, *Macaca*- und *Callithrix*-Sequenz an. CNSiPY = „Conserved non-coding sequence in *Pan troglodytes* Y". CNSiMmulY = „CNS in *Macaca mulatta* Y". CNSiCY = „CNS in *Callithrix jacchus* Y".

Homologe Sequenzblöcke	Hsap *DDX3Y*			Ptro *DDX3Y*			% Homologie
	BAC-Position	Position 2,35 kb Sequenz	bp Länge	BAC-Position	Position 2,34 kb Sequenz	bp Länge	
CNSiPY-1	52.646 - 55.190	1 - 2.347	2.347	61.314 - 58.980	1 - 2.335	2.335	97,9 (98,4)
Homologe Sequenzblöcke	**Hsap *DDX3Y***			**Mmul *DDX3Y***			**% Homologie**
	BAC-Position	**Position 2,35 kb Sequenz**	**bp Länge**	**BAC-Position**	**Position 2,33 kb Sequenz**	**bp Länge**	
CNSiMmulY-1	52.646 - 55.190	1 - 2.347	2.347	85.273 - 82.948	1 - 2.326	2.326	90,9 (91,7)
Homologe Sequenzblöcke	**Hsap *DDX3Y***			**Cjac *DDX3Y***			**% Homologie**
	BAC-Position	**Position 2,35 kb Sequenz**	**bp Länge**	**BAC-Position**	**Position 2,16 kb Sequenz**	**bp Länge**	
CNSiCY-1	52.646 - 52.892	1 - 247	247	53.768 - 53.542	1 - 227	227	81 (88,1)
CNSiCY-2	52.952 - 54.395	307 - 1.552	1.246	53.537 - 52.327	232 -1.442	1.211	80,3 (82,7)
CNSiCY-3	54.470 - 55.190	1.627- 2.347	721	52.307 - 51.606	1.462 - 2.163	702	74,5 (76,5)

In Tabelle 4-17 sind die kartierten CpG-Islands der sechs *DDX3Y/Ddx3y* Promoterregionen aufgelistet. In den Primaten ist jeweils ein CGI mit den relaxierten Parametern identifizierbar. Das Schimpansen *DDX3Y* CGI ist vergleichbar zum Human CGI (Abb. 4-50). In *Macaca* und *Callithrix* ist das CGI jeweils um über 100 Nukleotide verkürzt. Für keine der Primatenspezies konnte ein CGI mit den stringenten Parametern festgestellt werden.

Zwischen der Human und den Non-Human Primaten *DDX3Y* Promoterregionen konnten 101 konservierte TFBS festgestellt werden (Tab. 4-18). Diese Y-spezifischen Primaten-cTFBS sind an 60 nicht überlappenden Positionen lokalisiert. Für 58 der dazugehörigen TFs ist eine Expression in Hodengewebe feststellbar. 13 der Y-spezifischen Primaten-cTFBS wurden in Human auch als X-Y homologe TFBS erkannt (gelb unterlegte Felder in Tab. 4-18). Die X-Y

homologe TFBS für SRY in Human *DDX3Y* Position 417 (Tab. 4-10) ist innerhalb der Primaten, außer in *Macaca* (Deletion), ebenfalls konserviert. Der Human X-Y homologe TFBS SOX5 in Human *DDX3Y* Position 1.288 (*MSY2-1;* Tab. 4-10) erscheint in Non-Human Primaten vorzugsweise als DMRT- und nicht als SOX5-Bindestelle. In Human *DDX3Y* Position 1.395 (*MSY2-2*) erscheint die Bindestelle als SOX5-TFBS. Im Human X-Y TFBS-Vergleich scheint kein Unterschied bezüglich der SOX5- u. DMRT-Bindestelle zu bestehen, der Vergleich zu den Non-Human Primaten zeigt aber, daß im *DDX3Y* Promoter die Bindestelle in *MSY2-1* als DMRT und in *MSY2-2* als SOX5 Bindestelle konserviert vorliegt. Komplette Listen aller *DDX3Y* TFBSs sind auf der beigefügten Daten-CD hinterlegt (Tab. A-18 bis A-20).

Tab. 4-17: Vertikaler Vergleich der pot. CpG-Islands in den *DDX3Y*/*Ddx3y* Promoterregionen von Mensch, Schimpanse, Rhesusaffe, *Callithrix*, Rind und Maus.

Hsap *DDX3Y* = 2.545 bp, BAC 475I1 (BAC-Position 52.646 - 55.190). „CpG Island Searcher" ist so konfiguriert, daß benachbarte CGIs mit einer Lücke≤ 100 bp vereint werden. Werte in Klammern in Spalte 4 und 7 (O/E CpG) geben die Anzahl der gefundenen CpG-Dinukleotide an. --- = kein CpG-Island mit gewählten Parametern.

DDX3Y	**CpG Island Searcher** (500 bp, 55%, 0,65 O/E CpG)			**EMBOSS CpG Plot** (200 bp, 50%, 0,6 O/E CpG)		
	Position/ Länge	**%GC**	**CpG O/E**	**Position/ Länge**	**%GC**	**CpG O/E**
Hsap	---	---	---	2.150 – 2.486 (337 bp)	51,04	1,25 (25 CpGs)
Ptro	---	---	---	1.962 - 2.276 (315 bp)	52,70	1,37 (27 CpGs)
Mmul	---	---	---	2.054 - 2.267 (214 bp)	51,87	1,05 (13 CpGs)
Cjac	---	---	---	1.900 - 2.104 (205 bp)	52,20	1,24 (15 CpGs)
Btau	---	---	---	1) 1.694 - 1.912 (219 bp) 2) 2.073 - 2.486 (414 bp)	1) 55,71 2) 52,17	1) 0,71 (10 CpGs) 2) 1,05 (26 CpGs)
Mmus	---	---	---	---	---	---

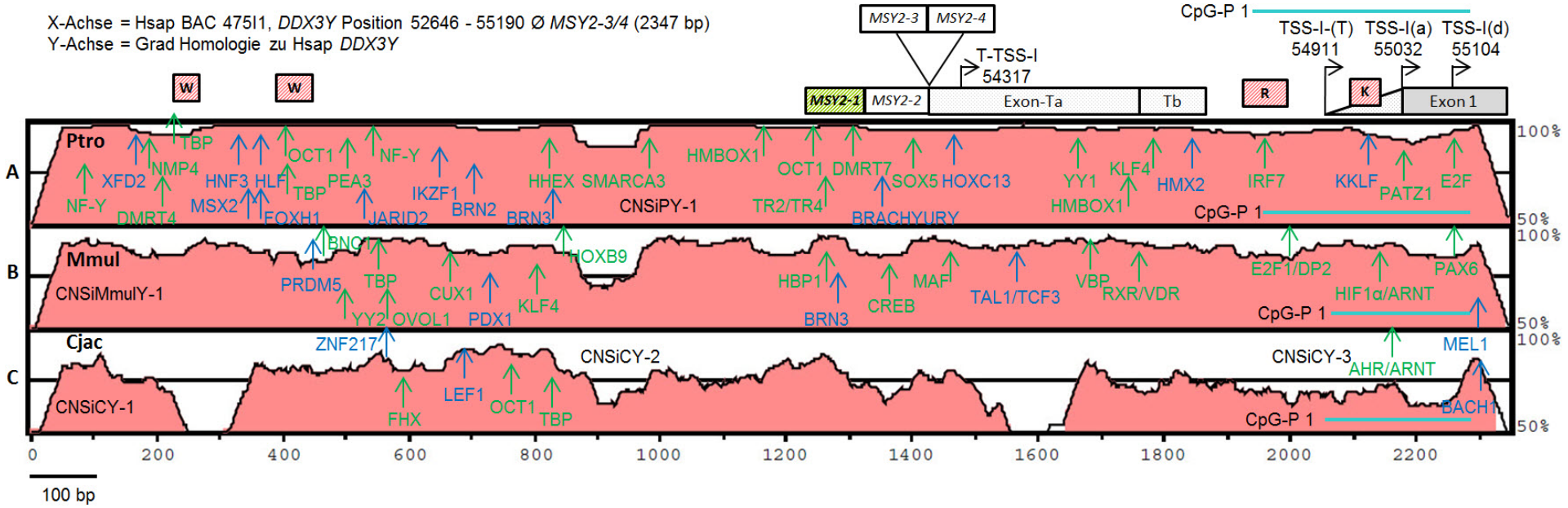


Abb. 4-50: mVISTA-, cTFBS- und CpG-Island-Vergleich der Human, Schimpanse, *Macaca* und *Callithrix DDX3Y* Promoterregionen.

Oberhalb des mVISTA-Plots ist die Human *DDX3Y* Promoter- und 5´UTR-Region schematisch dargestellt. Die X-Achse repräsentiert die Human *DDX3Y* Sequenz. Die Y-Achse gibt jeweils die Sequenzhomologie der verglichenen Primatenart an. **A** = Schimpanse, **B** = *Macaca*, **C** = *Callithrix*. Die konservierten Primaten Y-TFBS-Positionen sind mit blauen und grünen Pfeilen angegeben. Die Y-cTFBSs mit bekannter Expression in Hodengewebe sind grün markiert. Diese Y-cTFBS kommen in den *DDX3Y* Promoterregionen aller vier Primatenarten vor. Es werden nur die nicht überlappenden cTFBSs gezeigt. Bei überlappenden cTFBSs wird jeweils die TFBS mit höchstem Matrix-Wert angegeben. Beschreibung der TFs siehe Tab. 4-18. Die kartierten CpG-Islands sind in hellblau markiert (Tab. 4-17). Die Human Promoterelemente sind entsprechend Abb. 4-47 angegeben. *MSY2-1* ist grün schraffiert. Die *DDX3Y* Sequenztrakte sind rot schraffiert.

Tab. 4-18: Konservierte TFBS zwischen der Human und non-Human Primaten *DDX3Y* Promoterregion.

Die cTFBS werden für die Human *DDX3Y* Promoterregion angegeben (2.545 bp; BAC-Position in 475I1 52.646 - 55.190). Gelb unterlegte Felder in Spalte 2 markieren TFBSs, die in Tab. 4-10 bereits als Human X-Y homologe TFBS erkannt wurden. In Spalte 3 sind die TFBS-Positionen entsprechend der Lage in der 2,5 kb Promoterregion angegeben. In Spalte 4 ist die Lage der cTFBS in den konservierten Primatenregionen CNSiPY (P1), CNSiMmulY (M1) und CNSiCY (C1 - C3) angegeben. Spalte 5 gibt die Orientierung der TFBS an. „Core-Werte" in Spalte 6 geben Übereinstimmung zur TF-Core-Konsensussequenz an. Werte sind von 0 bis 1 angegeben, wobei 1 einer 100% Übereinstimmung zur Konsensussequenz entspricht. „Matrix-Werte" in Spalte 7 bewerten die Core-Sequenz und die umliegenden Nukleotide. „Opt-Werte" (Spalte 8) sind individuelle Grenzwerte für jede TF-Matrix und sollen nach Möglichkeit oberhalb von 0,85 liegen. Als Matrix-Bibliothek wurde nur die Vertebraten Matrix-Gruppe gewählt. In Spalte 9 ist angegeben, ob für die TFs eine Expression in Hodengewebe bekannt ist. Bei bekannter Hodenexpression ist Feld grün unterlegt. k. A. = keine Angabe zur Expression in Hodengewebe. Spalte 10 gibt ermittelte Bindesequenz in untersuchter Human Sequenz an. Die Sequenz kann revers-komplementär sein, wenn TF an Minus-Strang bindet.

Nr.	TF-Name	TF-Familie	Position	Lage CNS	Polarität *DDX3Y* Strang	Core-Wert	Matrix-Wert	Opt-Wert	Bekannte Hoden-expression	TF-Bindesequenz
1	NF-Y	„CCAAT-binding TF"; NFYA/HAP2 TF-Familie	86 - 100	P1, M1, C1	+	0,75	0,834	0,83	ja	cccaCCACtgcactc
2	XFD2	„Fork-head domain" TF-Familie	175 - 191	P1, M1, C1	+	1	0,912	0,89	k. A.	atacttTAAAcatacag
3	NMP4	Krueppel-C2H2-Zinkfinger TF-Familie	197 - 207	P1, M1, C1	-	1	0,978	0,97	ja	ggAAAAaaaac
	E2F	E2F/DP TF-Familie	199 - 215	P1, M1, C1	-	1	0,787	0,75	ja	tctcgggggGAAAaaaa
4	DMRT4	„Doublesex- and mab-3-related TF"	218 - 238	P1, M1, C1	-	1	0,797	0,79	ja	ataaatttaTACAgtcgttgc
5	BRN2	POU TF-Familie	224 - 242	P1, M1, C1	+	0,857	0,882	0,82	k. A.	actgtataaatTTATattg
	TBP	„TATA-binding Protein" TF-Familie	226 - 242	P1, M1, C1	-	1	0,925	0,9	ja	caataTAAAtttataca
6	HNF3	„Fork-head domain" TF-Familie	312 - 328	P1, M1, C2	-	0,971	0,986	0,98	k. A.	tagctgcAAATacaaac
7	PHOX2A	Paired-Homeobox TF-Familie	318 - 338	P1, M1, C2	-	1	0,879	0,87	k. A.	ttgccTAATttagctgcaaat
	NKX2-5	NK-2 Homeobox TF-Familie	320 - 338	P1, M1, C2	-	1	0,896	0,88	k. A.	ttgccTAATttagctgcaa
	GSH2	ANTP-Homeobox TF-Familie	321 - 339	P1, M1, C2	-	1	0,966	0,95	k. A.	gttgccTAATttagctgca
	MSX2	MSH-Homeobox TF-Familie	322 - 340	P1, M1, C2	-	1	0,977	0,95	k. A.	ggttgcCTAAtttagctgc
	RAX	Paired-Homeobox-Bicoid TF-Subfamilie	322 - 338	P1, M1, C2	-	1	0,921	0,89	k. A.	ttgccTAATttagctgc
8	HLF	PAR/bZIP TF-Familie	326 - 342	P1, M1, C2	+	0,836	0,87	0,84	k. A.	ctaaattagGCAAccac

9	FOXH1	„Fork-head domain“ TF-Familie	339 - 355	P1, M1, C2	+	1	0,916	0,81	k. A.	ccacttgtgTATTtgtc
10	FHX	„Fork-head domain“ TF-Familie	390 - 406	P1, M1, C2	-	1	0,85	0,83	ja	tttataATAAaatatgt
10	OCT1	POU TF-Familie	391 - 407	P1, M1, C2	+	1	0,865	0,85	ja	catattttATTAtaaat
10	BRN2	POU TF-Familie	396 - 414	P1, M1, C2	+	0,967	0,942	0,86	k. A.	ttTATTataaatattcttt
11	XFD2	„Fork-head domain“ TF-Familie	397 - 413	P1, M1, C2	+	1	0,933	0,89	k. A.	ttattaTAAAtattctt
	TBP	„TATA-binding Protein“ TF-Familie	398 - 414	P1, M1, C2	+	1	0,933	0,9	ja	tattaTAAAtattcttt
12	PRDM5	C2H2-Zinkfinger TF-Familie	431 - 459	P1, M1, C2	+	0,96	0,74	0,71	k. A.	aggtttgttAGGGgccagagggg tgtcat
13	BNC1	C2H2-Zinkfinger TF-Familie	450 - 468	P1, M1, C2	+	0,789	0,856	0,85	ja	ggggtgtcatTGTGcatca
14	PEA3	ETS TF-Familie	499 - 519	P1, M1, C2	+	1	0,947	0,94	ja	agaggaaAGGAaggtcaaaaa
15	YY2	YY TF-Familie	507 - 527	P1, M1, C2	-	1	0,972	0,96	ja	cttctCCATttttgaccttcc
16	JARID2	„AT-rich interactive domain“ TF-Familie	523 - 541	P1, M1, C2	-	1	0,896	0,89	k. A.	ataaaTTTAttggccttct
17	NF-Y	„CCAAT-binding TF“; NFYA/HAP2 TF-Familie	525 - 539	P1, M1, C2	+	1	0,94	0,92	ja	aaggCCAAtaaattt
18	EVI1	C2H2-Zinkfinger TF-Familie	533 - 549	P1, M1, C2	-	0,75	0,795	0,79	k. A.	gatatAAAAtaaattta
	TBP	„TATA-binding Protein“ TF-Familie	534 - 550	P1, M1, C2	-	1	0,95	0,9	ja	tgataTAAAataaattt
	CDX2	Caudal-Homeobox TF-Familie	535 - 553	P1, M1, C2	+	1	0,842	0,84	k. A.	aatttatTTTAtatcacgt
	HOXA9	Abdominal-B Homeobox TF-Familie	537 - 553	P1, M1, C2	-	1	0,858	0,85	ja	acgtgataTAAAataaa
	TEF	PAR/bZIP TF-Familie	539 - 555	P1, M1, C2	-	0,818	0,874	0,85	k. A.	acacgtgatATAAaata
19	ZNF217	Krueppel-C2H2-Zinkfinger TF-Familie	569 - 581	P1, M1, C2	+	1	0,913	0,9	k. A.	GAATggtaaaggt
20	OVOL1	OVO homolog-like TFs	570 - 584	P1, M1, C2	-	0,75	0,803	0,8	ja	gacaccTTTAccatt
21	FHX	„Fork-head domain“ TF-Familie	591 - 607	P1, M1, C2	-	1	0,831	0,83	ja	ttttttATAAcatattt
22	IKZF1	IKAROS-C2H2-Zinkfinger TF-Familie	657 - 669	P1, M1, C2	+	1	0,937	0,92	k. A.	ctctGGGAacact
23	CUX1	CUT-Homeobox TF-Familie	683 - 701	P1, M1, C2	+	1	0,983	0,94	ja	atgttcCAATcactttgaa

24	LEF1	TCF/LEF TF-Familie	690 - 706	P1, M1, C2	-	1	0,969	0,94	k. A.	aaacattCAAAgtgatt
	CHR	Cell cycle regulators: Cell cycle homology element	693 - 705	P1, M1, C2	+	1	0,947	0,92	k. A.	cactTTGAatgtt
	NANOG	NANAOG-Homeobox TF-Familie	693 - 711	P1, M1, C2	+	1	0,952	0,94	ja	cactttgAATGttttaatt
25	BRN2	POU TF-Familie	696 - 714	P1, M1, C2	+	0,841	0,846	0,82	k. A.	tttgaatgtttTAATttta
26	PDX1	ANTP-Homeobox TF-Familie	735 - 753	P1, M1, C2	-	0,826	0,76	0,74	k. A.	ctcaaaacaTTATcacaaa
27	OCT1	POU TF-Familie	760 - 776	P1, M1, C2	+	1	0,847	0,81	ja	ggtatttgAATTttagc
28	KLF4	Krueppel-C2H2-Zinkfinger TF-Familie	779 - 795	P1, M1, C2	-	1	0,991	0,96	ja	cttcAAAGggtttccta
	LEF1	TCF/LEF TF-Familie	783 - 799	P1, M1, C2	-	1	0,971	0,94	k. A.	cgtccttCAAAgggttt
29	LHX3	LIM-Homeobox TF-Familie	805 - 827	P1, M1, C2	+	1	0,89	0,82	ja	aaatctggatattTAATctggtt
	HHEX	Homeobox TF-Familie	807 - 825	P1, M1, C2	+	1	0,952	0,95	ja	atctggatattTAATctgg
30	BRN3	POU TF-Familie	808 - 826	P1, M1, C2	-	0,75	0,791	0,78	k. A.	accagattaAATAtccaga
29	TBP	„TATA-binding Protein“ TF-Familie	811 - 827	P1, M1, C2	+	0,75	0,79	0,78	ja	ggatattTAATctggtt
31	HOXB9	Abdominal-B Homeobox TF-Familie	851 - 867	P1, M1, C2	+	1	0,895	0,88	ja	agtgtttgTAAAaccac
32	SMARCA3	SNF2/RAD54 Helikasen	984 - 994	P1, M1, C2	-	1	0,986	0,98	ja	agttACTTttt
33	HMBOX1	Homeobox TF-Familie	1.174 - 1.190	P1, M1, C2	+	1	0,94	0,83	ja	tagcactgGTTAaccta
	HNF1	HNF1 Homeobox TF-Familie	1.179 - 1.195	P1, M1, C2	-	1	0,922	0,84	k. A.	caacctagGTTAaccag
34	OCT1	POU TF-Familie	1.220 - 1.236	***MSY2-1***	P1, M1, C2	+	1	0,802	0,77	ja	ccTATGccagtgtgcgt
35	TR2/TR4	„NR2-Nuclear Hormone Receptor“ TF-Familie	1.233 - 1.257		P1, M1, C2	+	1	0,782	0,77	ja	gcgttaagggcctaGGTCccccggt
36	HBP1	HMG-Box TFs	1.269 - 1.291		P1, M1, C2	-	1	0,906	0,86	ja	ttatatactgAATGaccctggtg
37	BRN3	POU TF-Familie	1.274 - 1.292		P1, M1, C2	+	0,75	0,802	0,78	k. A.	gggtcattcAGTAtataat
38	LHX3	LIM-Homeobox TF-Familie	1.276 - 1.298		P1, M1, C2	+	1	0,847	0,82	ja	gtcattcagtataTAATtgtttc
	SOX5	HMG-Box TFs	1.280 - 1.302		P1, M1, C2	-	1	0,985	0,87	ja	caatgaaaCAATtatatactgaa

38	EN1	ENGRAILED-Homeobox TF-Familie	1.282 - 1.300		P1, M1, C2	-	1	0,86	0,86	ja	atgaaacAATTatatactg
	NOBOX	Homeobox TF-Familie	1.282 - 1.300		P1, M1, C2	+	1	0,978	0,92	ja	cagtataTAATtgtttcat
	NKX1-2	NK1 Homeobox TF-Familie	1.284 - 1.300		P1, M1, C2	+	1	0,874	0,87	ja	gtatatAATTgtttcat
	NKX2-5	NK-2 Homeobox TF-Familie	1.284 - 1.302		P1, M1, C2	+	1	0,883	0,88	k. A.	gtataTAATtgtttcattg
	DMRT2	„Doublesex- and mab-3-related TF“	1.285 - 1.305		P1, M1, C2	-	0,75	0,76	0,75	ja	aaacaatgaAACAattatata
	DMRT7	„Doublesex- and mab-3-related TF“	1.287 - 1.307		P1, M1, C2	+	1	0,94	0,81	ja	tataATTGtttcattgtttcg
39	BRACHYURY	T-Box TF-Familie	1.356 - 1.376	***MSY2-2***	P1, M1, C2	+	1	0,682	0,66	k. A.	tccccttgtAGGTgtgttata
40	CREB	CREB-bZIP TF-Familie	1.375 - 1.395		P1, M1, C2	-	1	0,926	0,89	ja	catactgaaTGACgttgatta
41	DLX3	Distal-less Homeobox TF-Familie	1.387 - 1.405		P1, M1, C2	+	1	0,95	0,91	k. A.	ttcagtatgTAATtgtttc
	SOX5	HMG-Box TFs	1.395 - 1.417		P1, M1, C2	-	1	0,988	0,87	ja	tgaagaaaCAATgaaacaatta c
	HMX2	HMX-Homeobox TF-Familie	1.397 - 1.415		P1, M1, C2	+	0,752	0,849	0,81	k. A.	aattgtTTCAttgtttctt
42	POU2F3	POU TF-Familie	1.633 - 1.649		P1, M1, C3	+	1	0,847	0,8	k. A.	cgtATGCtcataaagcg
	MAF	bZIP-MAF TF-Familie	1.634 - 1.654		P1, M1, C3	+	1	0,902	0,88	ja	gtaTGCTcataaagcgcaggc
43	CDX2	CAUDAL-Homeobox TF-Familie	1.635 - 1.653		P1, M1, C3	-	1	0,858	0,85	k. A.	cctgcgctTTATgagcata
	HOXC13	Abdominal-B Homeobox TF-Familie	1.635 - 1.651		P1, M1, C3	+	1	0,92	0,83	k. A.	tatgctcaTAAAgcgca
44	TAL1/TCF3 Heterodimer	bHLH TF-Familie	1.768 - 1.788		P1, M1, C3	+	1	0,987	0,98	k. A.	ctttcctCAGGtgcagtcagg
	DEC2	bHLH TF-Familie	1.770 - 1.784		P1, M1, C3	-	0,871	0,973	0,96	k. A.	actgcaCCTGaggaa
45	YY1	YY TF-Familie	1.868 - 1.888		P1, M1, C3	-	1	0,957	0,94	ja	agtgaCCATtttacgtgagcg
	HOXA9/MEIS 1 Dimer	Abdominal-B Homeobox TF-Familie	1.872 - 1.886		P1, M1, C3	-	1	0,847	0,78	ja	TGACcattttacgtg
	NR1D1	„NR1-Nuclear Hormone Receptor“ TF-Familie	1.873 - 1.895		P1, M1, C3	+	1	0,897	0,88	ja	acgtaaaatgGTCActgcgcct a
	RXR/NR1H4 Heterodimer	„NR2/1-Nuclear Hormone Receptor“ TF-Familie	1.881 - 1.893		P1, M1, C3	+	0,797	0,805	0,8	ja	TGGTcactgcgcc
46	RXR/NR1I3 Heterodimer	„NR2/1-Nuclear Hormone Receptor“ TF-Familie	1.897 - 1.921		P1, M1, C3	-	0,792	0,761	0,75	ja	gataagGGTTacgggatcccag gtg
	VBP	PAR/bZIP Familie	1.901 - 1.917		P1, M1, C3	+	1	0,888	0,86	ja	tgggatcccGTAAccct

47	HMBOX1	Homeobox TF-Familie	1.940 - 1.956	P1, M1, C3	-	0,771	0,875	0,83	ja	tggaactaGTCAaaaaa
48	RXR/ PPAR	„NR2/1-Nuclear Hormone Receptor“ TF-Familie	1.963 - 1985	P1, M1, C3	+	1	0,723	0,69	ja	ctgaagcttgtcAAAGgtaggtg
49	RXR/VDR	„NR2/1-Nuclear Hormone Receptor“ TF-Familie	1.966 - 1.990	P1, M1, C3	+	1	0,824	0,75	ja	aagcttgtcaaaggtAGGTgacggc
48	KLF4	Krueppel-C2H2-Zinkfinger TF-Familie	1.971 - 1.987	P1, M1, C3	+	1	0,966	0,96	ja	tgtcAAAGgtaggtgac
50	HMX2	HMX-Homeobox TF-Familie	2.045 - 2.063	P1, M1, C3	+	1	0,904	0,83	k. A.	gagagagtgCTTAaggaag
51	IRF7	„Interferon Regulatory Factor“ TF-Familie	2.170 - 2.190	P1, M1, C3	+	0,937	0,891	0,86	ja	cctaGAATtcgaaggcattag
52	E2F1/DP2	E2F/DP TF-Familie	2.192 - 2.208	P1, M1, C3	+	1	0,816	0,78	ja	agagGGCGgggatagca
53	ZF9 (KLF6)	Krueppel-C2H2-Zinkfinger TF-Familie	2.319 - 2.341	P1, M1, C3	-	0,923	0,908	0,87	ja	gttggctCCACccctccaccacc
	KKLF (KLF15)	SP1-C2H2-Zinkfinger TF-Familie	2.321 - 2.337	P1, M1, C3	+	1	0,955	0,91	k. A.	tggtggaGGGGtggagc
	SP1	SP1-C2H2-Zinkfinger TF-Familie	2.326 - 2.340	P1, M1, C3	+	0,872	0,909	0,88	ja	gaggGGTGgagccaa
54	HIF1-alpha/ ARNT Dimer	bHLH TF-Familie	2.344 - 2.360	P1, M1, C3	-	1	0,936	0,91	ja	acgcacaaCGTGcgtaa
55	AHR/ARNT Heterodimer	binden an DRE (Dioxin Response Element) und XRE (Xenobiotic Response Element); bHLH TF-Familie	2.347 - 2.371	P1, M1, C3	+	1	0,967	0,95	ja	cgcacgttgtGCGTgctgtcgtata
56	PATZ1	Krueppel-C2H2-Zinkfinger TF-Familie	2.375 - 2.387	P1, M1, C3	+	1	0,921	0,88	ja	gtgggaGGGGcca
57	E2F	E2F/DP TF-Familie	2.477 - 2.493	P1, M1, C3	-	0,857	0,867	0,84	ja	gagaagcgaAAAAtcgg
58	PAX6	Paired-Homeobox TF-Familie	2.480 - 2.498	P1, M1, C3	+	0,755	0,773	0,75	ja	attttTCGCttctcttcag
59	MEL1	**C2H2-Zinkfinger TF-Familie**	2.493 - 2.509	P1, M1, C3	+	1	0,999	0,99	k. A.	cttcaggGATGagtcat
60	BACH1	bZIP-CNC TF-Familie	2.496 - 2.516	P1, M1, C3	+	1	0,982	0,82	k. A.	cagggaTGAGtcatgtggtgg
	MITF	bHLH TF-Familie	2.502 - 2.516	P1, M1, C3	+	1	0,95	0,92	k. A.	tgagtCATGtggtgg
	c-MYC/MAX	bHLH TF-Familie	2.504 - 2.516	P1, M1, C3	-	0,86	0,927	0,92	ja	ccaccaCATGact

Rind:

Zwischen der Human *DDX3Y* und der Rind *Ddx3y* Promoterregion konnten sechs konservierte Regionen identifiziert werden (Tab. 4-19). Eine Region CNSiBY-1 erfüllt die Bedingungen für ein CNS (100 bp u. 70% Sequenzhomologie). Die Region ist im Bereich des Human Exon-Ta und Exon-Tb lokalisiert (Abb. 4-51).

In der Rind *Ddx3y* Promoterregion konnten zwei CGIs kartiert werden (Tab. 4-17). Übertragen auf die Human *DDX3Y* Promoterkarte ist das erste CGI in Exon-Ta lokalisiert. Das zweite CGI erstreckt sich von Exon-Tb bis in das Exon1 hinein (Abb. 4-51).

Der Vergleich der TFBSs ergab 32 konservierte TFBSs zwischen Human *DDX3Y* und Rind *Ddx3y* an 19 nicht überlappenden Positionen (Tab. 4-20). Fünf der Human–Rind cTFBSs (Y) wurden bereits im Human X-Y Vergleich erkannt (gelb markiert in Tab. 4-20). Für 16 der Human–Rind cTFBSs (Y) ist eine Expression in Hodengewebe bekannt. Eine Auflistung aller Rind TFBS in der Rind *Ddx3y* Promoterregion ist auf der Daten-CD hinterlegt (Tab. A-21).

Tab. 4-19: Liste der mit mVISTA ermittelten konservierten Regionen zwischen der Human *DDX3Y* und Rind *Ddx3y* Promoterregion.

Spalte 2: Position der ermittelten Elemente im Human Y-Chr. BAC 475I1, bezogen auf die verkürzte Sequenz (ohne *MSY2-2/3/4* = 2.237 bp). In Klammern sind die Positionen der KRiBYs für die vollständige, 2,55 kb lange Human *DDX3Y* Promotersequenz angegeben. Spalte 3: Position der ermittelten Elemente in der 2.237 bp langen Human *DDX3Y* Sequenz (ohne *MSY2-2/3/4*). Spalte 4: Länge der ermittelten Elemente in der Human *DDX3Y* Promoterregion. Spalte 5: Position der ermittelten Elemente im Rind Y-BAC CH240-56A13 (rev.-komplementär). Spalte 6: Position der ermittelten Elemente in der 2.545 bp langen Rind *Ddx3y* Sequenz. Spalte 7: Länge der ermittelten Elemente in der Rind *Ddx3y* Promoterregion. Spalte 8: Sequenzidentität der Elemente in %. Werte in Klammern geben die Homologiewerte jeweils für die Rindersequenz an. KRiBY = „Konservierte Region in *Bos taurus* Y“. CNS ist wie in Abb. rosa markiert.

Homologe Sequenzblöcke	Hsap *DDX3Y*			Btau *Ddx3y*			% Homologie
	BAC-Position	Position 2,24 kb Sequenz	bp Länge	BAC-Position	Position 2,55 kb Sequenz	bp Länge	
KRiBY-1	53.221 - 53.279	576 - 634	59	141.713 - 141.638	579 - 654	76	69,5 (53,9)
KRiBY-2	53.349 - 53.414	704 - 769	66	141.598 - 141.527	694 - 765	72	65,2 (59,7)
KRiBY-3	53.875 - 53.954	1.230 - 1.309	80	140.704 - 140.616	1.588 - 1.676	89	71,3 (64)
CNSiBY-1	54.541 - 54.690	1.588 - 1.737 (1.896 - 2.045)	150	140.360 - 140.216	1.932 - 2.076	145	70 (71,7)
KRiBY-4	54.541 - 54.959	1.588 - 2.006 (1.896 - 2.314)	419	140.360 - 139.924	1.932 - 2.368	437	65,6 (62,9)
KRiBY-5	55.114 - 55.177	2.161 - 2.224 (2.469 - 2.532)	64	139.833 - 139.760	2.459 - 2.532	74	75 (65,8)

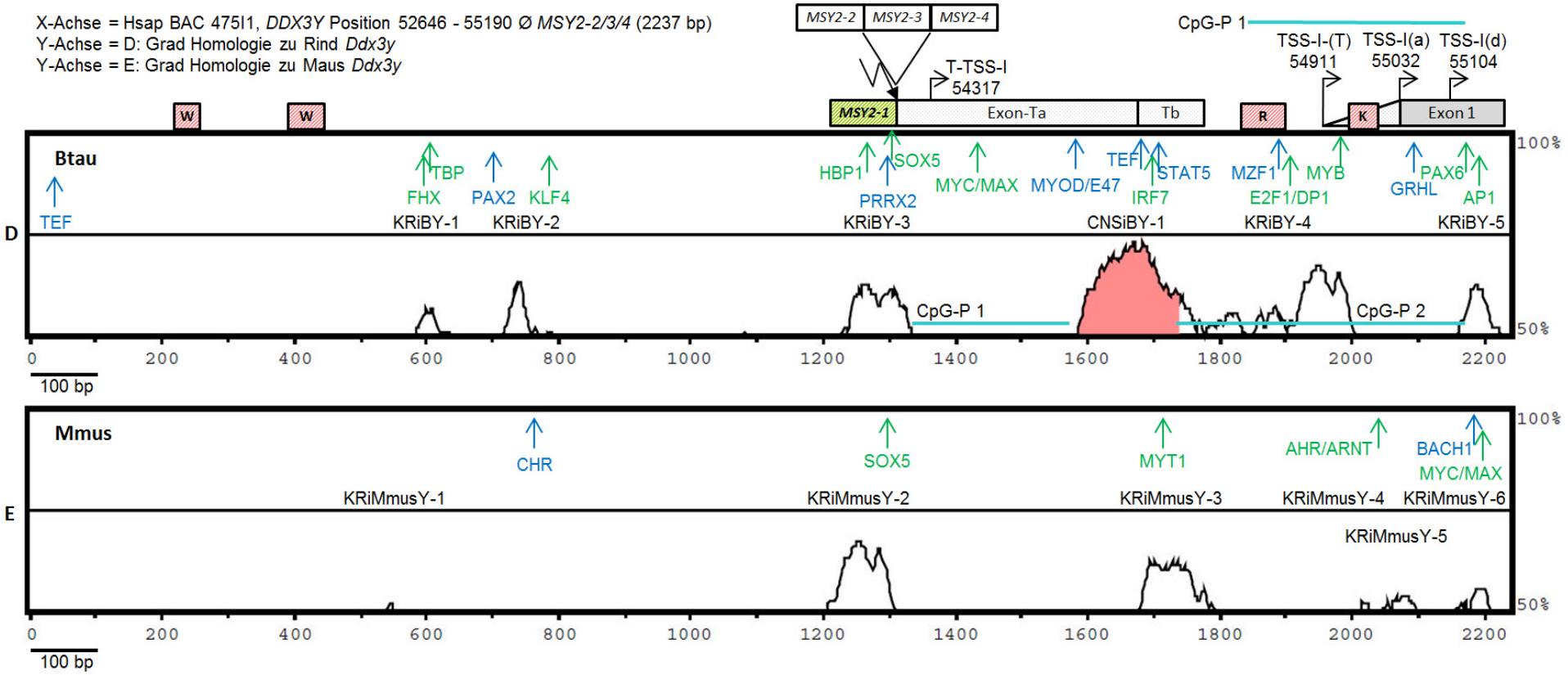


Abb. 4-51: mVISTA-, cTFBS- und CpG-Island-Vergleich der Human *DDX3Y* und Rind und Maus *Ddx3y* Promoterregionen.

D) Basislinie (X-Achse) ist Human *DDX3Y* (BAC 475I1, Position 52.646 – 55.190) ohne die *MSY2* Repeateinheiten 2, 3 und 4 (308 bp entfernt, Bruch durch Blitz angezeigt und als Einschub oberhalb gezeigt). Oberhalb des Plots ist Human *DDX3Y* Promoter- und 5´UTR-Region schematisch dargestellt (siehe Abb. 4-47). *DDX3Y* ATG ist an BAC-Position 55.146 (hier rel. Position 2.193). Y-Achse gibt den Grad an Konserviertheit in Rind *Ddx3y* Sequenz wieder. Fünf konservierte Regionen und ein CNS konnten identifiziert werden (siehe Tab. 13). KRiBY = „Konservierte Region in *Bos taurus* Y". „Conserved non-coding sequences" CNS ist in rosa dargestellt.

E) Basislinie (X-Achse) ist Human *DDX3Y* (BAC 475I1, Position 52.646 – 55.190) ohne die *MSY2* Repeateinheiten 2, 3 und 4. Y-Achse gibt den Grad an Konserviertheit in Maus *Ddx3y* Sequenz wieder. KRiMmusY = „Konservierte Region in *Mus musculus* Y".

Die konservierten Y-TFBS-Positionen sind mit blauen und grünen Pfeilen angegeben. Die Y-cTFBSs mit bekannter Expression in Hodengewebe sind grün markiert.

Tab. 4-20: Konservierte TFBS in Human *DDX3Y* und Rind *Ddx3y*.

Human *DDX3Y* = 2.545 bp, BAC 475I1 (Position 52.646 - 55.190). Rind *Ddx3y* = 2.545 bp, BAC-Klon BAC CH240-56A13 (Position 142.291 - 139.747, rev.-komplementär). In Spalte 2 können für Rind *Ddx3y* andere TF-Bezeichnungen auftreten, da „DiAlign TF" cTFBS für TF-Familien angibt. Identische TF-Core- und Matrix-Werte zwischen Human *DDX3Y* und Rind *Ddx3y* werden nicht gesondert aufgeführt, veränderte Angaben für Rind *Ddx3y* sind in Klammern angegeben. Gelb unterlegte Felder in Spalte 2 markieren TFBSs, die in Tab. 4-10 bereits als Human X-Y homologe TFBS erkannt wurden. In Spalte 3 sind die TFBS-Positionen entsprechend der Lage in der 2,5 kb Promoterregion angegeben. In Spalte 4 ist die Lage der TFBS in den konservierten Regionen KRiBY („Konservierte Region in *Bos taurus* Y") angegeben. Spalte 5 gibt die Orientierung der TFBS an. Der „Core-Wert" in Spalte 6 gibt die Übereinstimmung zur TF-Core-Konsensussequenz an. Werte sind von 0 bis 1 angegeben, wobei 1 einer 100% Übereinstimmung zur Konsensussequenz entspricht. „Matrix-Werte" in Spalte 7 bewerten die Core-Sequenz und die umliegenden Nukleotide. „Opt-Werte" (Spalte 8) sind individuelle Grenzwerte für jede TF-Matrix und sollen nach Möglichkeit oberhalb von 0,85 liegen. Als Matrix-Bibliothek wurde nur die Vertebraten Matrix-Gruppe gewählt. In Spalte 9 ist angegeben, ob für TF eine Expression in Hodengewebe bekannt ist. Bei bekannter Hodenexpression ist Feld grün unterlegt. k. A. = keine Angabe zur Expression in Hodengewebe. Spalte 10 gibt ermittelte Bindesequenz in untersuchter Sequenz an. Die Sequenz kann revers-komplementär sein, wenn TF an Minus-Strang bindet.

Nr.	TF-Name	TF-Familie	Position	Lage KRiBY		Polarität *DDX3Y* Strang	Core-Wert	Matrix-Wert	Opt-Wert	Bekannte Hoden-expression	TF-Bindesequenz
1	TEF (VBP)	PAR/bZIP TF-Familie	41 - 57 (180 - 196)			-	1	0,863 (0,886)	0,85 (0,86)	k. A.	caggttcagGTAAttca (tattttaaaGTAActca)
2	FHX	„Fork-head domain" TF-Familie	584 - 600 (601 - 617)	**Y-1**		+	0,818	0,841 (0,83)	0,83	ja	cgacacAAAAtatgtta (agtcacAAAAcacttgc)
3	TBP	„TATA-binding Protein" TF-Familie	596 - 612 (615 - 631)	**Y-1**		+	1	0,97 (0,944)	0,9	ja	tgttaTAAAaaaggaaa (tgctaTAAAgaaggctg)
4	PAX2	Paired-Homeobox TF-Familie	697 - 719 (688 - 710)	**Y-2**		+	0,789 (1)	0,801 (0,818)	0,78	k. A.	ttgaatgttttaattttAAATtc (tgaatatttccattttgAAACag)
5	KLF4 (BKLF)	Krueppel-C2H2-Zinkfinger TF-Familie	779 - 795 (776 - 792)			-	1	0,991 (0,981)	0,96 (0,95)	ja	cttcAAAGggtttccta (ctgGGGTggggtctcct)
6	HBP1	HMG-Box TFs	1269 - 1291 (1635 - 1657)	***MSY2***	**Y-3**	-	1	0,906 (0,867)	0,86	ja	ttatatactgAATGaccctggtg (ttatacgttgAATGacctgttga)
7	ISL1	LIM-Homeobox TF-Familie	1280 - 1302 (1646 - 1668)		**Y-3**	+	1	0,829 (0,827)	0,82	k. A.	ttcagtataTAATtgtttcattg (ttcaacgtaTAATtgttcaattg)
	DLX3	Distal-less Homeobox TF-Familie	1280 - 1298 (1646 - 1664)		**Y-3**	+	1	0,959	0,91	k. A.	ttcagtataTAATtgtttc (ttcaacgtaTAATtgttca)
	SOX5	HMG-Box TFs	1280 -1302 (1646 - 1668)		**Y-3**	-	1	0,985 (0,983)	0,87	ja	caatgaaaCAATtatatactgaa (caattgaaCAATtatacgttgaa)
	NOBOX	Homeobox TF-Familie	1282 - 1300 (1648 - 1666)		**Y-3**	+	1	0,978	0,92	ja	cagtataTAATtgtttcat (caacgtaTAATtgttcaat)
	NKX2-5	NK-2 Homeobox TF-Familie	1284 - 1302 (1650 - 1668)		**Y-3**	+	1	0,883	0,88	k. A.	gtataTAATtgtttcattg (acgtaTAATtgttcaattg)
	PRRX2 (S8)	Paired-Homeobox TF-Familie	1284 - 1304 (1650 - 1670		**Y-3**	+	1	0,995	(0,97)	k. A.	gtataTAATtgtttcattgtt (acgtaTAATtgttcaattgtt)
	RAX	Paired-Homeobox-Bicoid TF-Subfamilie	1284 - 1300 (1650 - 1666)		**Y-3**	+	1	0,891 (0,899)	0,89	k. A.	gtataTAATtgtttcat (acgtaTAATtgttcaat)

7	DLX2	Distal-less Homeobox TF-Familie	1285 - 1303 (1651 - 1669)		**Y-3**	-	1	0,942 (0,946)	0,92	k. A.	acaatgaaacAATTatata (acaattgaacAATTatacg)
	DMRT7	„Doublesex- and mab-3-related TF“	1287 - 1307 (1653 - 1673)		**Y-3**	+	1	0,94 (0,813)	0,81	ja	tataATTGtttcattgtttcg (tataATTGttcaattgttacg)
8	SOX5	HMG-Box TFs	1288 - 1310 (1654 - 1676)		**Y-3**	-	1	0,988 (0,987)	0,87	ja	tcacgaaaCAATgaaacaattat (tgacgtaaCAATtgaacaattat)
9	ARNT	bHLH/PAS TF-Familie	1739 - 1755 (1772 - 1788)			+	1	0,975 (0,949)	0,89	ja	tggtcacaCGTGctgcg (cgccagcaCGTGctgcg)
	DEC2 (HELT)	bHLH TF-Familie	1740 - 1754 (1773 - 1787)			-	1	0,983 (0,928)	0,96 (0,91)	k. A.	gcagcaCGTGtgacc (gcagCACGtgctggc)
	c-MYC/MAX Heterodimer	bHLH TF-Familie	1742 - 1754 (1774 - 1786)			-	1	0,991 (0,993)	(0,92)	ja	gcagcaCGTGtga (gcagcaCGTGctg)
10	MYOD/E47 Heterodimer	bHLH TF-Familie	1892 - 1908 (1928 - 1944)	**Y-4 CNSiBY-1**		-	1	0,962 (0,996)	0,93 (0,92)	k. A.	ggatcccaGGTGttagg (tgaccGCAGgtgtccag)
11	E4BP4	bZIP-NFIL3 TF-Subfamilie	2003 - 2023 (2034 - 2054)	**Y-4 CNSiBY-1**		-	1	0,811 (0,81)	0,8	k. A.	aggaagtttcGTAAaattttt (ctgaatttctGTAAaatgtct)
	TEF	PAR/bZIP TF-Familie	2006 - 2022 (2037 - 2053)	**Y-4 CNSiBY-1**		-	1	0,874 (0,888)	0,85	k. A.	ggaagtttcGTAAaatt (tgaatttctGTAAaatg)
12	IRF7	„Interferon Regulatory Factor“ TF-Familie	2011 - 2031 (2042 - 2062)	**Y-4 CNSiBY-1**		-	0,937	0,88 (0,898)	(0,86)	ja	ctgaGAATaggaagtttcgta (ccacGAATctgaatttctgta)
13	STAT5	STAT TF-Familie	2019 - 2037 (2050 - 2068)	**Y-4 CNSiBY-1**		-	1 (0,945)	0,977 (0,906)	(0,89)	k. A.	ttacTTCTgagaataggaa (ttagTTCCacgaatctgaa)
14	MZF1	Krueppel C2H2-Zinkfinger TF-Familie	2197 - 2207 (2242 - 2252)	**Y-4**		+	1	0,992 (1)	(0,99)	k. A.	gcGGGGatagc (gtGGGGagagg)
15	E2F1/DP1 (E2F4/DP2)	E2F/DP TF-Familie; Zellzyklus G1 zu S	2207 - 2223 (2250 - 2266)	**Y-4**		-	1 (0,847)	0,793 (0,822)	0,71 (0,76)	ja	aattcGCGCgatgtttg (aactcGCGAgaactcct)
16	OVOL1	OVO homolog-like TFs	2285 - 2299 (2339 - 2353)	**Y-4**		-	1	0,942 (0,919)	0,8	ja	gctaccGTTActtct (tctaccGTTActtct)
	MYB	„myb-like transcriptional regulators“	2288 - 2300 (2342 - 2354)	**Y-4**		+	1	0,991	0,9	ja	agtAACGgtagcg (agtAACGgtagat)
17	GRHL	GRH/CP2 Grainyhead Subfamilie	2401 - 2413 (2409 - 2421)			-	1	0,886 (0,874)	0,86	k. A.	cgcactGGTTagc (ctcacaGGTTggt)
18	PAX6	Paired-Homeobox TF-Familie	2480 - 2498 (2480 - 2498)	**Y-5**		+	0,755	0,773 (0,754)	0,75	ja	attttTCGCttctcttcag (agttcGCGCctttcttcag)
19	BACH1	bZIP-CNC TF-Familie	2494 - 2514 (2494 - 2514)	**Y-5**		-	0,75	0,858 (0,867)	0,82	k. A.	accacaTGACtcatccctgaa (gcctcaTGACtcatgcctgaa)
	AP1	bZIP-JUN TF-Familie	2500 - 2510 (2500 - 2510)	**Y-5**		-	0,905	0,973	0,94	ja	catgACTCatc (catgACTCatg)

Maus:

Zwischen den Human und Maus *DDX3Y/Ddx3y* Promoterregionen konnten sechs konservierte Sequenzabschnitte festgestellt werden (Tab.4-21). Dabei erreicht keine der kartierten Abschnitte die für ein CNS geforderten Grenzwerte. Der mVISTA-Plot ist in Abb. 4-51 zu sehen. Ein CpG-Island konnte in der Maus *Ddx3y* Promoterregion nicht festgestellt werden (Tab. 4-17).

Es konnten 7 konservierte TFBSs an 6 nicht überlappenden Positionen lokalisiert werden (Tab. 4-22). Fünf der cTFBSs zeigen eine Expression in Hodengewebe und fünf der Human–Maus cTFBSs (Y) wurden bereits im Human X-Y Vergleich erkannt (gelb markiert in Tab. 4-22).

Im cTFBS(Y)-Vergleich aller sechs Spezies, konnten nur die SOX5-TFBS, im *MSY2* Repeat, und die BACH1-TFBS, überlappend mit dem Translationsstartcodon ATG, als hoch-konservierte TFBS in allen sechs untersuchten Säugerspezies identifiziert werden. Diese zwei TFBS wurden auch in Human als X-Y konserviert erkannt. Eine Auflistung aller Maus TFBS in der Maus *Ddx3y* Promoterregion ist auf der Daten-CD hinterlegt (Tab. A-22).

Tab. 4-21: Liste der mit mVISTA ermittelten konservierten Regionen zwischen der Human *DDX3Y* und Maus *Ddx3y* Promoterregion.

Spalte 2: Position der ermittelten Elemente im Human Y-Chr. BAC 475I1, bezogen auf die verkürzte Sequenz (ohne *MSY2-2/3/4* = 2.237 bp). Spalte 3: Position der ermittelten Elemente in der 2.237 bp langen Human *DDX3Y* Sequenz (ohne *MSY2-2/3/4*). In Klammern sind die Positionen der KRiMmusYs für die vollständige, 2,55 kb lange Human *DDX3Y* Promotersequenz angegeben. Spalte 4: Länge der ermittelten Elemente in der Human *DDX3Y* Promoterregion. Spalte 5: Position der ermittelten Elemente im Maus Y-BAC RP24-208N6 (2.545 bp, Position 96.244 - 93.700, rev.-komplementär). Spalte 6: Position der ermittelten Elemente in der 2545 bp langen Maus *Ddx3y* Sequenz. Spalte 7: Länge der ermittelten Elemente in der Maus *Ddx3y* Promoterregion. Spalte 8: Sequenzidentität der Elemente in %. Werte in Klammern geben die Homologiewerte jeweils für die Maussequenz an. KRiMmusY = „Konservierte Region in *Mus musculus* Y“.

Homologe Sequenzblöcke	Hsap *DDX3Y*			Mmus *Ddx3y*			% Homologie
	BAC-Position	Position 2,24 kb Sequenz	bp Länge	BAC-Position	Position 2,55 kb Sequenz	bp Länge	
KRiMmusY-1	53.181 - 53.196	536 - 551	16	95.621 - 95.604	624 - 641	18	62,5 (55,6)
KRiMmusY-2	53.854 - 53.954	1.209 - 1.309	101	94.372 - 94.276	1.873 - 1.969	97	63,6 (70,1)
KRiMmusY-3	54.628 - 54.745	1.75 - 1.792 (1.983 - 2.100)	118	94.168 - 94.051	2.077 - 2.194	118	58,5
KRiMmusY-4	54.962 - 54.980	2.009 - 2.027 (2.317 - 2.335)	19	93.915 - 93.892	2.330 - 2.353	24	63,2 (50)
KRiMmusY-5	54.996 - 55.050	2.043 - 2.097 (2.351 - 2.405)	55	93.878 - 93.830	2.367 - 2.415	49	52,7 (59,2)
KRiMmusY-6	55.108 - 55.167	2.155 - 2.214 (2.463 - 2.522)	60	93.787 - 93.723	2.458 - 2.522	65	68,3 (63,1)

Tab. 4-22: Konservierte TFBS in Human *DDX3Y* und Maus *Ddx3y*.

Human *DDX3Y* = 2.545 bp, BAC-Klon 475I1 (Position 52.646 - 55.190). Für Maus = 2.545 bp, BAC-Klon RP24-208N6 (Position 96.244 - 93.700, rev.-komplementär). In Spalte 2 können für Maus *Ddx3y* andere TF-Bezeichnungen auftreten, da „DiAlign TF" cTFBS für TF-Familien angibt. Identische TF-Core- und Matrix-Werte zwischen Human *DDX3Y* und Maus *Ddx3y* werden nicht gesondert aufgeführt, veränderte Angaben für Maus *Ddx3y* sind in Klammern angegeben. Gelb unterlegte Felder in Spalte 2 markieren TFBSs, die in Tab. 4-10 bereits als Human X-Y homologe TFBS erkannt wurden. In Spalte 3 sind die TFBS-Positionen entsprechend der Lage in der 2,5 kb Promoterregion angegeben. In Spalte 4 ist die Lage der TFBS in den konservierten Regionen KRiMmusY („Konservierte Region in *Mus musculus* Y") angegeben. Spalte 5 gibt die Orientierung der TFBS an. „Core-Werte" in Spalte 6 gibt Übereinstimmung zur Konsensus-TF-Core Sequenz an. Werte sind von 0 bis 1 angegeben, wobei 1 einer 100% Übereinstimmung zur Konsensussequenz entspricht. „Matrix-Werte" in Spalte 7 bewerten die Core-Sequenz und die umliegenden Nukleotide. „Opt-Werte" (Spalte 8) sind individuelle Grenzwerte für jede TF-Matrix und sollen nach Möglichkeit oberhalb von 0,85 liegen. Als Matrix-Bibliothek wurde nur die Vertebraten Matrix-Gruppe gewählt. In Spalte 9 ist angegeben, ob für TF eine Expression in Hodengewebe bekannt ist. Bei bekannter Hodenexpression ist Feld grün unterlegt. k. A. = keine Angabe zur Expression in Hodengewebe. Spalte 10 gibt ermittelte Bindesequenz in untersuchter Sequenz an. Die Sequenz ist revers-komplementär angegeben, wenn TF an Minus-Strang bindet.

Nr.	TF Name	TF Familie	Position	Lage KRi-MmusY	Polarität *DDX3Y* Strang	Core-Wert	Matrix-Wert	Opt-Wert	Bekannte Hoden-expression	TF-Bindesequenz
1	CHR	Cell cycle regulators: Cell cycle homology element	761 - 773 (1.531 - 1.543)		+	1	0,986 (0,957)	0,92	k. A.	gtatTTGAatttt (caatTTGAatatt)
2	SOX5	HMG-Box TFs	1.288 - 1.310 (1.951 - 1.967)	**Y-2 (*MSY2*)**	-	1	0,988 (0,993)	0,87	ja	tcacgaaaCAATgaaacaattat (agaaaaCAATaaaaatt)
3	MYT1 (Myelin TF 1)	C2HC Zinkfingerprotein	2.011 - 2.023 (2.105 - 2.117)	**Y-3**	-	1 (0,750)	0,882 (0,799)	0,88 (0,75)	ja	aggAAGTttcgta (agaAATTttacta)
4	AHR/ARNT Heterodimer	binden an DRE (Dioxin Response Element) und XRE (Xenobiotic Response Element); bHLH TF-Familie	2.347 - 2.371 (2.363 - 2.387)		+	1	0,967 (0,938)	0,95	ja	cgcacgttgtGCGTgctgtcgtata (cgtgcgtcttgCGTGatgacgcaaa)
5	BACH1 (BACH2)	bZIP-CNC TF-Familie	2.496 - 2.516 (2.496 - 2.516)	**Y-6**	+	1	0,982 (0,971)	0,82 (0,89)	k. A.	cagggaTGAGtcatgtggtgg (cagcaaTGAGtcaagtggcag)
	AP1 (JUN)	bZIP-Jun TF-Subfamilie	2.500 - 2.510 (2.499 - 2.511)	**Y-6**	-	0,905	0,973 (0,951)	0,94	ja	catgACTCatc (acttgACTCattg)
6	c-MYC/ MAX (USF1 & 2)	bHLH TF-Familie	2.504 - 2.516 (2.504 - 2.516)	**Y-6**	-	0,86 (0,851)	0,927 (0,912)	0,92 (0,9)	ja	ccaccaCATGact (ctgccaCTTGact)

DDX3X:

Für die *DDX3X*/*Ddx3x* mVISTA-Analyse wurden folgende Sequenzen verwendet: Schimpanse X-Chromosomale Sequenz[45], *Macaca* X-Chromosomale Sequenz[46], *Callithrix* BAC-Klon „CH259-42H12“[47], Rind X-Chromosomale Sequenz[48], Maus BAC-Klon „RP23-158L19“[49].
Im Unterschied zum *DDX3Y/Ddx3y* mVISTA-Plot konnte für den Vergleich der *DDX3X/Ddx3x* Promoterregionen ein einheitlicher Plot aller sechs Spezies ohne Qualitätsverlust erstellt werden.

Der mVISTA-Plot zwischen der Human und Schimpansen *DDX3X* Promoterregion ergab eine durchgehend konservierte Sequenz mit einer Sequenzhomologie von 98% (Tab. 4-23, Abb. 4-52). Das Human–*Macaca* Alignment ist durch zwei kurze indels (Insertion/Deletion) unterbrochen. Die Sequenzhomologien für *Macaca* liegen zwischen 92 und 95%. Für *Callithrix* konnten zwei lange konservierte Regionen kartiert werden, die im Bereich der Pyrimidin-reichen Sequenz (Y-Trakt) im Human *DDX3X* Promoter unterbrochen werden. Die Sequenzhomologien liegen bei 90 und 80%. Für die Rind *Ddx3x* Promoterregion konnten fünf konservierte Sequenzabschnitte mit Sequenzhomologien zwischen 68 und 80% identifiziert werden. In der potentiellen Maus *Ddx3x* Promoterregion konnten acht, meist kürzere Sequenzabschnitte, mit Sequenzhomologien zwischen 62 und 82% kartiert werden (Abb. 4-52).

Die CPG-Island Analyse ergab für alle *DDX3X/Ddx3x* Promoterregionen CGIs mit den stringenten, als auch den relaxierten CGI-Parametern (Tab. 4-24). Das Muster innerhalb der Primaten stellt sich sehr einheitlich dar. Es sind jeweils ein langes, durchgehendes CGI mit den stringenten Parametern und zwei kürzere CGIs mit den relaxierten Parametern kartierbar (Abb. 4-52). Das CpG-P2 ist jeweils bei *Macaca* und *Callithrix* ausgedehnter, als bei Mensch und Schimpanse, und erstreckt sich 5´ bis in den Bereich der Human T/B-TSS-II Position. In der Rind Promoterregion konnte mit den stringenten Parametern ein langes CGI, vergleichbar zum Human CGI, kartiert werden. Mit den relaxierten Parametern konnten drei CGIs festgestellt werden. Das Primaten CGI-P1 ist bei Rind in zwei kurze CGIs unterteilt. Das Maus CGI mit stringenten Parametern ist deutlich kürzer, als in den anderen Säugern, und reicht 5´ bis in den Bereich des Human S-Traktes. Das relaxierte Maus CGI ist 5´ etwas länger als das Human CGI-P2. Insgesamt ist das proximale Human *DDX3X* CGI-P2 in allen untersuchten Spezies kartierbar und stellt somit ein konserviertes Promoterfunktionselement dar.

---

45 *DDX3X* Ptro: X-Chromosomale Pos. 41.564.565 - 41.565.440 (876 bp, Ensembl Vers. CHIMP2.1) und BAC-Klon „CH251-3A20“, Position 1 - 1.640, 1.640 bp + 876 = 2.516 bp gesamt.

46 *DDX3X* Mmul: X-Chromosomale Pos. 39.099.681 - 39.102.191 (2511 bp; Ensembl Vers. MMUL1).

47 *DDX3X* Cjac: BAC-Pos. 99.190 - 96.717, 2.472 bp, revers-komplementär.

48 *DDX3X* Btau: X-Chromosomale Pos. 82.419 - 84.963 („Un.004.81“, 2.545 bp, Ensembl Btau4.0).

49 *DDX3X* Mmus: BAC-Pos. 88.039 - 90.583, 2.545 bp.

Tab. 4-23: Liste der mit mVISTA ermittelten konservierten Sequenzabschnitte zwischen den Human, Non-Human Primaten, Rind und Maus *DDX3X/Ddx3x* Promoterregionen.

Spalte 2: Position der ermittelten Elemente im Human X-Chr. BAC RP13-13A3. Spalte 3: Position der ermittelten Elemente in der 2.545 bp langen Human *DDX3X* Sequenz. Spalte 4: Länge der ermittelten Elemente in der Human *DDX3X* Promoterregion. Spalte 5: Chromosomale und BAC-Positionen der ermittelten Elemente in den Schimpanse, *Macaca, Callithrix*, Rind und Maus *DDX3X/Ddx3x* Promoterregionen. Spalte 6: Position der ermittelten Elemente in den 2.516 bp langen Schimpanse, 2.511 bp langen *Macaca*, 2.474 bp langen *Callithrix*, und den 2.545 bp langen Rind und Maus *DDX3X/Ddx3x* Sequenzen. Spalte 7: Länge der ermittelten Elemente in den *DDX3X/Ddx3x* Promoterregionen. Spalte 8: Sequenzidentität der Elemente in %. Werte in Klammern geben die Homologiewerte jeweils für die Schimpansen-, *Macaca*-, *Callithrix*, Rind und Maus-Sequenz an. CNSiPX = „Conserved non-coding sequence in *Pan troglodytes* X“. CNSiMmulX = „CNS in *Macaca mulatta* X“. CNSiCX = „CNS in *Callithrix jacchus* X“. KRiBX = Konservierte Region in *Bos taurus* X. KRiMmusX = Konservierte Region in *Mus musculus* X. CNSs sind wie in Abb. 4-52 rosa markiert.

Homologe Sequenzblöcke	Hsap *DDX3X*			Ptro *DDX3X*			% Homologie
	BAC-Position	Position 2,55 kb Sequenz	bp Länge	X-Chr.-Position	Position 2,52 kb Sequenz	bp Länge	
CNSiPX-1	28.230 - 30.774	1 - 2545	2545	41.564.565 - 41.567.753	1 - 2516	2516	98,1 (99,2)
Homologe Sequenzblöcke	**Hsap *DDX3X***			**Mmul *DDX3X***			**% Homologie**
	BAC-Position	**Position 2,55 kb Sequenz**	**bp Länge**	**X-Chr.-Position**	**Position 2,51 kb Sequenz**	**bp Länge**	
CNSiMmulX-1	28.230 - 30.213	1 - 1984	1984	39.099.681 - 39.101.633	1 - 1953	1953	91,8 (93,2)
KRiMmulX-1	30.214 - 30.252	1985 - 2023	39	39.101.655 - 39.101.693	1975 - 2013	39	94,9
CNSiMmulX-2	30.279 - 30.774	2050 - 2545	496	39.101.694 - 39.102.191	2014 - 2511	498	94,8 (94,4)
Homologe Sequenzblöcke	**Hsap *DDX3X***			**Cjac *DDX3X***			**% Homologie**
	BAC-Position	**Position 2,55 kb Sequenz**	**bp Länge**	**BAC-Position**	**Position 2,47 kb Sequenz**	**bp Länge**	
CNSiCX-1	28.233 - 29.867	4 - 1638	1635	99.187 - 97.545	4 - 1646	1643	90,3 (89,9)
CNSiCX-2	29.913 - 30.774	1684 - 2545	862	97.544 - 96.717	1647 - 2474	801	80 (86,1)
Homologe Sequenzblöcke	**Hsap *DDX3X***			**Btau *Ddx3x***			**% Homologie**
	BAC-Position	**Position 2,55 kb Sequenz**	**bp Länge**	**BAC-Position**	**Position 2,55 kb Sequenz**	**bp Länge**	
KRiBX-1	28.235 - 28.365	6 - 136	131	84.779 - 84.919	45 - 185	141	74 (68,8)
KRiBX-2	28.388 - 29.485	159 - 1256	1098	83.687 - 84.725	239 - 1277	1039	68,3 (72,2)
KRiBX-3	29.518 - 29.701	1289 - 1472	184	83.458 - 83.640	1324 - 1506	183	70,7 (71)
KRiBX-4	29.715 - 30.161	1486 - 1932	447	82.982 - 83.421	1543 - 1982	440	68,9 (70)
CNSiBX-1	30.282 - 30.774	2053 - 2545	492	82.419 - 82.914	2050 - 2545	496	79,5 (78,8)
Homologe Sequenzblöcke	**Hsap *DDX3X***			**Mmus *Ddx3x***			**% Homologie**
	BAC-Position	**Position 2,55 kb Sequenz**	**bp Länge**	**BAC-Position**	**Position 2,55 kb Sequenz**	**bp Länge**	
CNSiMmusX-1	28.942 - 29.142	713 - 913	201	88.984 - 89.181	946 - 1143	198	75,1 (76,3)
KRiMmusX-1	29.149 - 29.196	919 - 967	48	89214 - 89.252	1176 - 1214	39	64,6 (79,5)

KRiMmusX-2	29.269 - 29.359	1040 - 1130	91	89.321 - 89.414	1283 - 1376	94	63,7 (61,7)
KRiMmusX-3	29.599 - 29.684	1370 - 1455	86	89.607 - 89.692	1569 - 1654	86	62,8
KRiMmusX-4	29.843 - 29.904	1614 - 1675	62	89.778 - 89.842	1740 - 1804	65	64,5 (61,5)
KRiMmusX-5	29.940 - 30.048	1711 - 1819	109	89.860 - 89.965	1822 - 1927	106	64,2 (66)
KRiMmusX-6	30.122 - 30.160	1893 - 1931	39	89.998 - 90.039	1960 - 2001	42	82,1 (76,2)
CNSiMmusX-2	30.435 - 30.774	2197 - 2545	340	90.240 - 90.583	2196 - 2545	344	75,9 (75)

Tab. 4-24: Vertikaler Vergleich der pot. CpG-Islands in den *DDX3X*/*Ddx3x* Promoterregionen von Mensch, Schimpanse, Rhesusaffe, *Callithrix*, Rind und Maus.

Für Hsap *DDX3X* = 2.545 bp, BAC RP13-13A3 (Position 28.230 - 30.774). „CpG Island Searcher" ist so konfiguriert, daß benachbarte CGIs mit einer Lücke≤ 100 bp vereint werden. Werte in Klammern in Spalte 4 und 7 (O/E CpG) geben die Anzahl der gefundenen CpG-Dinukleotide an.

DDX3X	**CpG Island Searcher** (500 bp, 55%, 0,65 O/E CpG)			**EMBOSS CpG Plot** (200 bp, 50%, 0,6 O/E CpG)		
	Position/ Länge	**%GC**	**CpG O/E**	**Position/ Länge**	**%GC**	**CpG O/E**
Hsap	782 – 2545 (1764 bp)	59,9	0,853 (135 CpGs)	1) 1120 – 1915 (796 bp) 2) 2238 – 2488 (251 bp)	1) 67,46 2) 60,96	1) 0,82 (71 CpGs) 2) 1,10 (25 CpGs)
Ptro	777 – 2516 (1740 bp)	60,2	0,849 (134 CPGs)	1) 1122 – 1915 (794 bp) 2) 2218 - 2459 (242 bp)	1) 67,76 2) 61,98	1) 0,82 (72 CpGs) 2) 1,15 (26 CpGs)
Mmul	754 – 2511 (1758 bp)	59,4	0,82 (127 CpGs)	1) 1199 – 1887 (689 bp) 2) 2010 - 2454 (445 bp)	1) 66,33 2) 59,10	1) 0,79 (58 CpGs) 2) 1,04 (39 CpGs)
Cjac	821 – 2474 (1654 bp)	59,1	0,826 (119 CpGs)	1) 1058 - 1862 (805 bp) 2) 1976 – 2417 (442 bp)	1) 65,47 2) 57,92	1) 0,84 (71 CpGs) 2) 0,99 (35 CpGs)
Btau	886 – 2545 (1660 bp)	58,3	0,774 (109 CpGs)	1) 1318 - 1556 (239 bp) 2) 1565 - 1844 (280 bp) 3) 2057 - 2488 (432 bp)	1) 66,53 2) 63,57 3) 57,41	1) 0,89 (23 CpGs) 2) 0,91 (17 CpGs) 3) 1,16 (40 CpGs)
Mmus	1809 – 2545 (737 bp)	57,8	0,764 (47 CpGs)	2174 - 2488 (315 bp)	59,68	1,00 (28 CpGs)

In Tabelle 4-25 sind die 31 TFBS aufgeführt, die in allen untersuchten *DDX3X/Ddx3x* Promoterregionen konserviert zu finden sind. Für die 31 cTFBSs konnten 22 nicht überlappende Positionen identifiziert werden. Für 16 der 31 cTFBSs ist eine Expression in Hodengewebe bekannt. Acht der X-spezifischen Säuger cTFBS wurden in Human auch als X-Y homologe TFBS erkannt (gelb unterlegte Felder in Tab. 4-25). Die X-Y homologe TFBS für SRY in Human *DDX3X* Position 533 (Tab. 4-10) ist innerhalb der Säuger, außer in Maus, ebenfalls konserviert. Im Vergleich zur *DDX3Y* Promoterregion ist insgesamt eine wesentlich höhere Anzahl an konservierten TFBS erhalten geblieben. Vollständige Listen aller *DDX3X/Ddx3x* TFBSs sind auf der beigefügten Daten-CD hinterlegt (Tab. A-23 bis A-27).

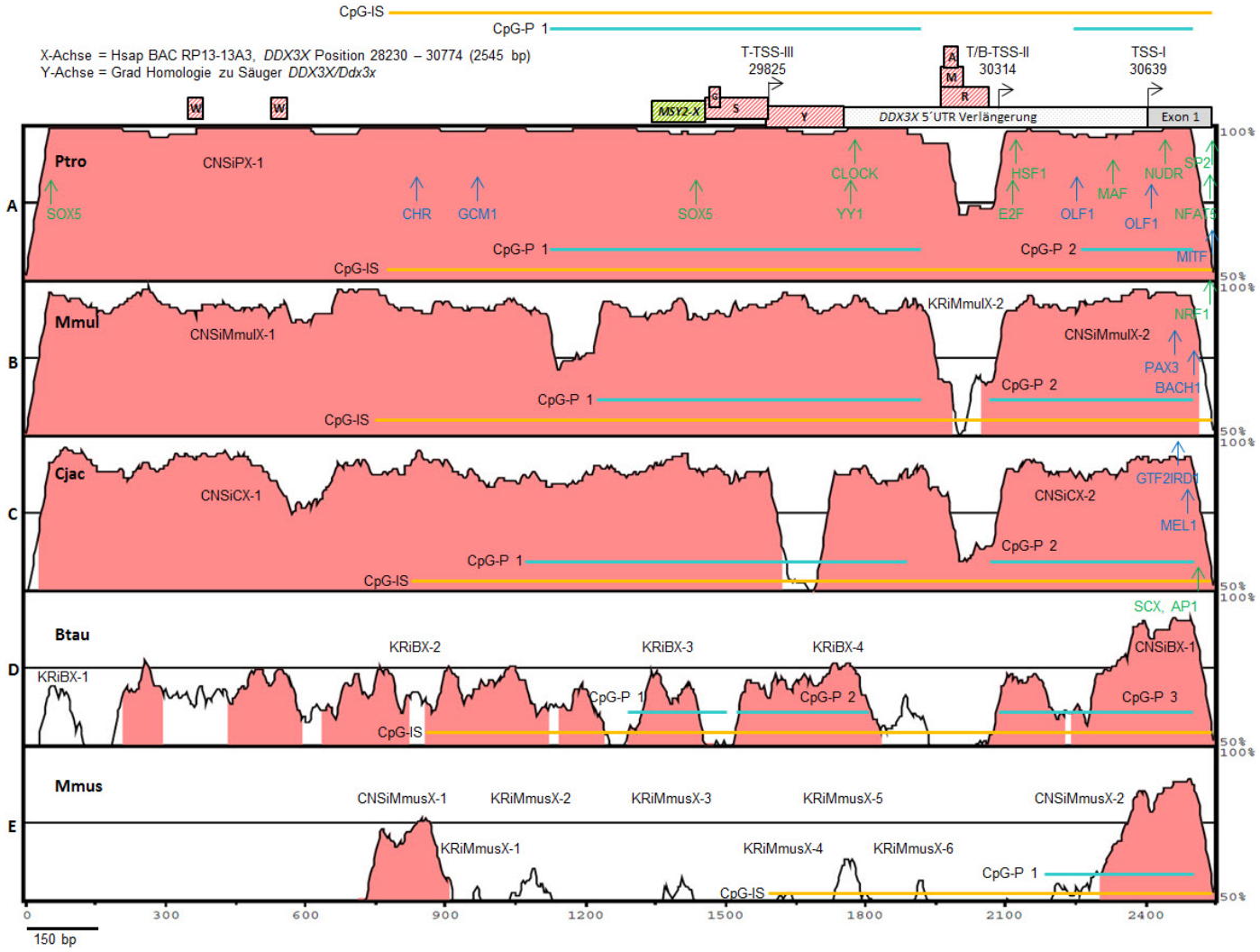


Abb. 4-52: mVISTA-Vergleich der Human, Schimpanse, *Macaca*, *Callithrix*, Rind und Maus *DDX3X/ Ddx3x* Promoterregionen.

Oberhalb des mVISTA-Plots ist die Human *DDX3X* Promoter- und 5´UTR-Region schematisch dargestellt. Die X-Achse repräsentiert die Human *DDX3X* Sequenz. Die Y-Achse gibt jeweils die Sequenz-homologie der verglichenen Säugerspezies an. A = Schimpanse, B = *Macaca*, C = *Callithrix*, D = Rind, E = Maus. Die konservierten Primaten X-TFBS-Positionen sind mit blauen und grünen Pfeilen angegeben. Die X-cTFBSs mit bekannter Expression in Hodengewebe sind grün markiert. Diese X-cTFBS kommen in den *DDX3Y* Promoterregionen aller sechs unter-suchter Säugerarten vor. Es werden nur die nicht überlappenden X-cTFBSs gezeigt. Bei überlappenden X-cTFBSs wird jeweils die TFBS mit höchstem Matrix-Wert angegeben. Beschreibung der TFs siehe Tab. 4-25. Die kartierten CpG-Islands sind in gelb (CpG Island Searcher) und hellblau (EMBOSS CpG Plot) markiert (Tab. 4-24). Die Human Promoterelemente sind entsprechend Abb. 4-47 angegeben. *MSY2-X* ist grün schraffiert. Die *DDX3X* Sequenztrakte sind rot schraffiert.

Tab. 4-25: Konservierte TFBS zwischen der Human *DDX3X* und den Non-Human Primaten, Rind und Maus *DDX3X/Ddx3x* Promoterregionen.

Human *DDX3X* = 2.545 bp, BAC RP13-13A3 (Position). In Spalte 3 sind die TFBS-Positionen entsprechend der Lage in der 2,55 kb Human *DDX3X* Promotersequenz. In Spalte 4 ist die Lage der TFBS in den konservierten Sequenzabschnitten angegeben. Spalte 5 gibt die Orientierung der TFBS an. „Core-Werte" in Spalte 6 gibt Übereinstimmung zur TF-Core-Konsensussequenz an. Werte sind von 0 bis 1 angegeben, wobei 1 einer 100% Übereinstimmung zur Konsensussequenz entspricht. „Matrix-Werte" in Spalte 7 bewerten die Core-Sequenz und die umliegenden Nukleotide. „Opt-Werte" (Spalte 8) sind individuelle Grenzwerte für jede TF-Matrix und sollen nach Möglichkeit oberhalb von 0,85 liegen. Als Matrix-Bibliothek wurde nur die Vertebraten Matrix-Gruppe gewählt. In Spalte 9 ist angegeben, ob für TF eine Expression in Hodengewebe bekannt ist. Bei bekannter Hoden-expression ist Feld grün unterlegt. k. A. = keine Angabe zur Expression in Hodengewebe. Spalte 10 gibt ermittelte Bindesequenz in der Human *DDX3X* Promotersequenz an. Die Sequenz kann revers-komplementär sein, wenn TF an Minus-Strang bindet.

Nr.	TF-Name	TF-Familie	Position	Lage CNS	*DDX3X* Strang	Core-Wert	Matrix-Wert	Opt-Wert	Bekannte Hoden-expression	TF-Bindesequenz
1	SOX5	HMG-Box TFs	50 - 72	P1, Mmul1, C1, B1	-	1	0,872	0,87	ja	gatttctaCAATgccccttggag
2	CHR	Cell cycle regulators: Cell cycle homology element	870 - 882	P1, Mmul1, C1, B2, Mmus1	+	1	0,993	0,92	k. A.	ctatTTGAattct
3	GCM1 (hGCMa)	Chorion-spezifischer TF	943 - 953	P1, Mmul1, C1, B2, Mmus1	-	1	0,918	0,85	k. A.	gaCCCTcagcc
4	SOX5/30	HMG-Box TFs	1425 - 1447	P1, Mmul1, C1, B3, Mmus3 (***MSY2-X***)	-	1	0,982	0,87	ja	tgagggaaCAATgagacagacac
5	YY1	YY TF-Familie	1763 - 1783	P1, Mmul1, C2, B4, Mmus5	-	1	0,978	0,94	ja	ggtggCCATtttgtgaagcag
6	XBP1	CREB-bZIP TF-Familie	1794 - 1814	P1, Mmul1, C2, B4, Mmus5	-	1	0,881	0,88	k. A.	ggcaggccACGTgaccgtagg
	CLOCK/ BMAL1 (ARNTL)	bHLH/PAS TF-Familie	1796 - 1812	P1, Mmul1, C2, B4, Mmus5	+	1	0,979	0,84	ja	tacggtCACGtggcctg
	DEC2 (bHLHE41)	bHLH TF-Familie	1997 - 1811	P1, Mmul1, C2, B4, Mmus5	-	1	0,978	0,96	k. A.	aggccaCGTGaccgt
7	E2F/DP1 Heterodimer	E2F/DP TF-Familie; Zellzyklus G1 zu S	2108 - 2124	P1, Mmul2, C2, B5, Mmus6	-	1	0,954	0,84	ja	agatctCGCGagaactt
8	HSF1	Heat-shock TF-Familie	2112 - 2136	P1, Mmul2, C2, B5, Mmus6	-	1	0,802	0,75	ja	tcggagttctCGAGatctcgcgaga
9	OLF1	G-protein coupled receptor 1 family	2250 - 2272	P1, Mmul2, C2, B5, Mmus6	+	1	0,883	0,88	k. A.	aggaaaTCCCttgagcttagacc
10	HIF1-alpha/ ARNT Dimer	bHLH TF-Familie	2341 - 2357	P1, Mmul2, C2, B5, Mmus6	+	1	0,928	0,91	ja	acgttgtaCGTGctgac
	MAF	bZIP-MAF TF-Familie	2344 - 2364	P1, Mmul2, C2, B5, Mmus6	+	1	0,993	0,97	ja	ttgtacgtGCTGacgtagccg
	CREB	CREB-bZIP TF-Familie	2345 - 2365	P1, Mmul2, C2, B5, Mmus6	+	1	0,971	0,89	ja	tgtacgtgcTGACgtagccgg

10	MAFA	bZIP-MAF TF-Familie	2348 - 2368	P1, Mmul2, C2, B5, Mmus6	+	1	0,959	0,92	k. A.	acgTGCTgacgtagccggctt
	CREB1	CREB-bZIP TF-Familie	2348 - 2368	P1, Mmul2, C2, B5, Mmus6	+	1	0,962	0,86	ja	acgtgcTGACgtagccggctt
	E4F	**C2H2-Zinkfinger TF-Familie**	2353 - 2365	P1, Mmul2, C2, B5, Mmus6	+	1	0,915	0,82	k. A.	ctgACGTagccgg
11	OLF1	„G-protein coupled receptor 1“ TF-Familie	2417 - 2439	P1, Mmul2, C2, B5, Mmus6	+	1	0,829	0,82	k. A.	cagttcTCCCgtgagagggcctt
12	NUDR (DEAF1)	MYND-Zinkfinger TF-Familie	2464 - 2482	P1, Mmul2, C2, B5, Mmus6	+	0,762	0,74	0,73	ja	gcaGCGGaagactccgagt
13	PAX3	Paired-Homeobox TF-Familie	2484 - 2502	P1, Mmul2, C2, B5, Mmus6	+	0,78	0,771	0,76	k. A.	cTCGGtactcttcagggat
14	**GTF2IRD1 (BEN)**	TFII-I Familie	2487 - 2501	P1, Mmul2, C2, B5, Mmus6	+	1	0,836	0,8	k. A.	ggtaCTCTtcaggga
15	MEL1	**C2H2-Zinkfinger TF-Familie**	2493 - 2509	P1, Mmul2, C2, B5, Mmus6	+	1	0,999	0,99	k. A.	cttcaggGATGagtcat
16	BACH1	bZIP-CNC TF-Familie	2494 - 2514	P1, Mmul2, C2, B5, Mmus6	-	0,75	0,858	0,82	k. A.	gccacaTGACtcatccctgaa
	BACH1	bZIP-CNC TF-Familie	2496 - 2516	P1, Mmul2, C2, B5, Mmus6	+	1	0,982	0,82	k. A.	cagggaTGAGtcatgtggcag
17	SCX	Scleraxis; TWIST-bHLH TF-Subfamilie	2500 - 2520	P1, Mmul2, C2, B5, Mmus6	+	1	0,934	0,91	ja	gatgagtcatgTGGCagtgga
18	AP1 (JUN)	bZIP-Jun TF-Subfamilie	2500 - 2510	P1, Mmul2, C2, B5, Mmus6	+	1	1	1	ja	gatGAGTcatg
19	MITF	bHLH TF-Familie	2502 - 2516	P1, Mmul2, C2, B5, Mmus6	+	1	0,949	0,92	k. A.	tgagtCATGtggcag
	USF1/2	bHLH TF-Familie	2504 - 2516	P1, Mmul2, C2, B5, Mmus6	-	0,915	0,929	0,9	ja	ctgcCACAtgact
20	NFAT5	„Nuclear factor of activated T-cells“	2515 - 2533	P1, Mmul2, C2, B5, Mmus6	+	1	0,878	0,87	ja	agtGGAAaatgcgctcggg
21	NRF1	NRF1/EWG Familie	2522 - 2538	P1, Mmul2, C2, B5, Mmus6	+	1	0,784	0,78	ja	aatGCGCtcgggctgga
22	SP2	SP1 C2H2-Zinkfinger TF-Familie	2528 - 2542	P1, Mmul2, C2, B5, Mmus6	+	1	0,858	0,8	ja	ctcgggctGGACcag

4.8. Unterschiedliche Gonosomenevolution im Bereich der *DDX3Y* und *DDX3X* Genloci

Unter dem Eindruck der deutlichen Veränderungen der Human *DDX3Y* und *DDX3X* Genloci in ihren Promoterbereichen, wurden die beiden Genloci, in einem 50 kb großen Genomfenster, mit dem Genomatix Genom-Annotations-Programm „ElDorado“ untersucht. Der Fokus der Analyse lag dabei auf der Kartierung von Sequenzelementen der Chromatin-Organisation, sogenannten S/MARs („Scaffold/Matrix Attachment Regions“), sowie von repetitiven DNA-Elementen der *Alu*-Familie in Bezug auf die *DDX3Y* und *DDX3X* Exonstrukturen (Abb. 4-53). Für den *DDX3Y* Locus wurden 50 kb der genomischen Y-Chromosom-Sequenz NC_000024.9 (Position 14.998.955 - 15.048.954; GRCh37.p2), für *DDX3X* 50 kb der genomischen X-Chromosom-Sequenz NC_000023.10 (Position 41.175.840 - 41.225.839) verwendet.

4.8.1. *DDX3Y* und *DDX3X* Exon- und Intronstrukturen

Die *DDX3Y* und *DDX3X* Exonstrukturen sind sehr homolog (Abb. 4-53, Tab. A-4, A-10; Kamp, 2001). Beide Gene verfügen über 17 Exone, die in den kodierenden Sequenzen (CDS) eine Homologie von 88,3% aufweisen (Tab. A-4, A-10). Die Homologie der Aminosäuresequenzen beträgt 92,4%. Die Exonlängen weichen nur in Exon1, je nach TSS, und in Exon17 stärker voneinander ab. Dagegen weist die Hälfte der Introns einen Längenunterschied von 50, oder mehr Nukleotiden auf. Die Distanzen zwischen Startcodon ATG (*DDX3Y* Pos. 15.016.848; *DDX3X* Pos. 41.193.506) und Stopcodon TGA (*DDX3Y* Pos. 15.030.032; *DDX3X* Pos. 41.206.970) sind mit 13.185 bp (Y) und 13.465 bp (X) sehr konserviert.

Für *DDX3Y* konnten drei Polyadenylierungssignale (PAS) in Exon17 festgestellt werden (Rauschendorf et al., 2011). Die zwei proximalen, internen *DDX3Y* PAS1 (ATTAAA; Pos. 15.030.255) und PAS2 (AATAAA; Pos. 15.030.420) werden in hodenspezifischen Transkripten genutzt. Die ubiquitären *DDX3Y* Transkripte enden in der distalen, konstitutiven PAS3 (AATAAA; Pos. 15.032.372). Für *DDX3X* konnten fünf PAS in Exon17 identifizierte werden. Die proximale, interne *DDX3X* PAS1 (AATAAA; Pos. 41.207.303) ist homolog zu *DDX3Y* PAS2 und wird ebenfalls in hodenspezifischen Transkripten genutzt. Die distalen *DDX3X* PAS2 (ATTAAA; Pos. 41.209.429) und PAS3 (AATAAA) sind überlappend und werden nicht gewebespezifisch exprimiert. *DDX3X* PAS3 ist homolog zu *DDX3Y* PAS3. 74 bp stromabwärts zu *DDX3X* PAS2 ist eine weitere ubiquitär genutzte PAS, *DDX3X* PAS4 (AATAAA; Pos. 41.209.503) genannt, lokalisiert. Die ubiquitär genutzten *DDX3Y* und *DDX3X* PAS entsprechen dem kanonischen Polyadenylierungssignal „AATAAA“, welches in über 70% der Humangene genutzt wird (Sheets et al., 1990; Zhao et al., 1999; Tian et al., 2005).

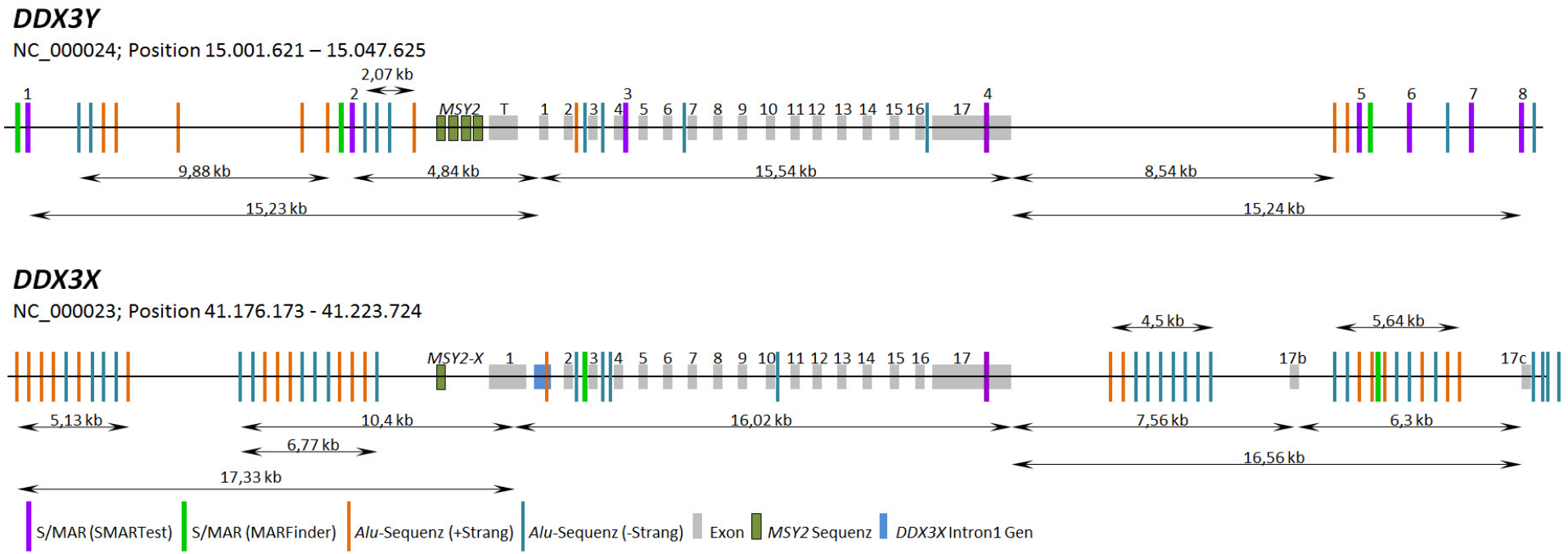


Abb. 4-53: Schematische Darstellung der Human Gonosomen im Bereich der *DDX3Y* und *DDX3X* Genloci.

Es wurde jeweils eine 50 kb lange Sequenz untersucht. Gezeigt werden fünf unterschiedliche Elemente. S/MAR = „Scaffold/Matrix Attachment Region“ (lila SMARTest; hellgrün MARFinder), repetitive DNA-Sequenzen der *Alu*-Familie (orange +Strang; blau -Strang), die Exone der *DDX3Y* und *DDX3X* Gene (grau), die *MSY2* Minisatellitensequenz (grün) und ein zusätzliches Gen im *DDX3X* Intron1 (hellblau). Die potentiellen S/MARs (SMARTest) im DX3Y Genlocus sind entsprechend der Tab. 4-26 nummeriert. Gezeigtes Schema gibt keine realistischen Größenverhältnisse und Abstände wieder. *DDX3Y* ATG = 15.016.848 entspricht BAC 475I1 Position 55.146. *DDX3X* ATG = 41.193.506 entspricht BAC RP13-13A3 Position 30.730. Es wurden nur repetitive Elemente der *Alu*-Familie in die Analyse einbezogen.

Die hodenspezifische *DDX3Y* PAS1 entspricht der Sequenz „ATTAAA“, der häufigsten PAS-Variante mit einem Einzelbasenaustausch (Tian et al., 2005). Die Erkennung und Polyadenylierung der hexameren PAS-Varianten ist generell geringer als für das kanonische Signal (Sheets et al., 1990). Die Distanzen zwischen ATG und konstitutivem 3´UTR Ende (*DDX3Y* Pos. 15.032.390; *DDX3X* Pos. 41.209.523) sind mit 15,54 kb für *DDX3Y* und 16,02 kb für *DDX3X* ebenfalls sehr stabil geblieben.

Stromabwärts des konstitutiven *DDX3X* Exon17 sind zwei weitere 3´UTR-Exone, *DDX3X* Exon 17b (Pos. 41.217.086 - 41.217.253) und 17c (Pos. 41.223.527 - 41.223.725) genannt, lokalisiert (Dickemann, 2004). Diese werden hodenspezifisch für *DDX3X* an eine interne Splicestelle in Exon17, Exon17a (Pos. 41.207.169) genannt, gesplict und generieren dadurch *DDX3X* Transkripte mit alternativen 3´UTR-Längen[50] (Dickemann, 2004). *DDX3X* Exon17b befindet sich 9,9 kb stromabwärts der Splicestelle in Exon17a und Exon17c ist nochmals 6,3 kb stromabwärts von 17b lokalisiert (Abb. 4-53). Die dazugehörige *DDX3X* PAS5 (AATAAA; Pos. 41.223.699) ist in Exon 17c lokalisiert. Mit Start in *DDX3X* T-TSS-III (Pos. 41.192.601) und Ende in Exon17c entsteht somit ein Primärtranskript mit einer Länge von 31.125 bp. Das Vorhandensein mehrerer alternativer Polyadenylierungssignale und somit die Entstehung unterschiedlicher 3´UTR-Varianten, wie in *DDX3Y* und *DDX3X*, kommt in über 50% der Säugergene vor (Iseli et al., 2002; Lee et al., 2008). Die alternativen Polyadenylierungsmuster sind von funktionellem Interesse, da sie z.B. als Biomarker für unterschiedliche Zelltypen, Zellstatus, Differenzierung und Proliferation dienen können (Ji & Tian, 2009; Ji et al., 2009).

DDX3X 3´UTR Splicingmuster:

Die *DDX3X* Exon17a/b/c Exonstruktur konnte in der genomischen Sequenz auch in Schimpanse, *Macaca* und *Callithrix* identifiziert werden. Die Exon-Intron-Grenzen wurden durch die kanonischen Splicing-Konsensussequenzen (5´-GT/3´-AG) festgelegt (Padgett et al., 1986). Bei *Callithrix* ist das 3´Splicesignal vor Exon17c mutiert. Experimentell wurde das *DDX3X* 3´UTR Exon17abc-Splicemuster bisher nur Hodengewebe von Mensch und Schimpanse nachgewiesen (Abb. 4-54). In Abb. 4-54 fällt auf, daß das Schimpansen PCR-Amplifikat größer ist, als das Human PCR-Produkt mit 618 bp. Die nachfolgende Sequenzanalyse ergab für *Pan* ein weiteres *DDX3X* Exon, 17d genannt (Ptro BAC CH251-3A20, Position 28.067 - 28.113), zwischen Exon17b und 17c. Das 47 bp lange Exon 17d ist beim Schimpansen in einer 1200 bp langen genomischen Region lokalisiert, die im Humangenom nicht existiert. Mit dem Programm „RepeatMasker“ konnten in diesem 1200 bp langen genomischen Intervall drei zusätzliche *Alu*-Elemente kartiert werden.

---

[50] Eine nicht charakterisierte und 5´ unvollständige *DDX3X* cDNA mit Exon17a/b/c Splicemuster ist unter GenBank Acc.No. HSAF000983.1 aufgeführt (Lahn & Page, 1997).

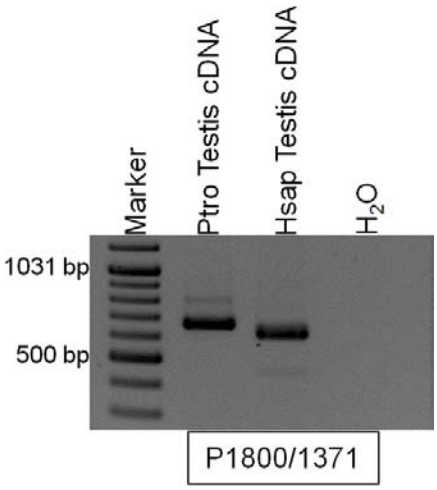


Abb. 4-54: RT-PCR Analyse des *DDX3X* Exon17 Splicemusters im Hodengewebe von Schimpanse und Human.
Erwartetes Human *DDX3X* PCR-Produkt = 618 bp. Der Forwardprimer P1800 ist überlappend in Exon16/17 (BAC-Pos. 43924) lokalisiert. Die Primerposition von P1371 ist im 3´Ende von Exon17c (PAC RP1-169I5, Pos. 709). Ptro = Schimpanse, Hsap = Mensch. 35 PCR Zyklen. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

Diese Schimpansen *DDX3X* 3´UTR entspricht also dem Muster Exon17abdc. Vom Javaneraffen *Macaca fascicularis* existiert in der NCBI-Datenbank ein cDNA-Klon aus Hodengewebe (GenBank Acc.No. AB168192.1) der ebenfalls das Splicemuster Exon17abdc aufweist.

4.8.2. S/MAR-Sequenzen und *Alu*-Elemente

„Scaffold/Matrix Attachment Regions“ (S/MARs):

Eine S/MAR-Konsensussequenz konnte bisher nicht identifiziert werden, was die Vorhersage von S/MARs erschwert. Unabhängig von einer Konsensussequenz konnten S/MAR-Eigenschaften, wie hoher (A,T)-Gehalt von über 70%, ~200 bis 3000 bp Länge, (A,T)-Trakte, Topoisomerase II-Erkennungssequenzen, gebogene DNA-Konformationen, „Nucleosome Excluding Sequences“ (NES) und „Inverted Repeats“ beschrieben werden (Cockerill & Garrard, 1986a, 1986b; Kay & Bode, 1994; Wang et al., 2010). Die Tendenz von S/MARs zur Strangtrennung geht von einem, für die Strangtrennung kritischen Sequenzelement „AATATATTT“ aus, dem „core unwinding element“, welches in einer BUR („base-unpairing region“) genannten Sequenz-region vorkommt (Bode et al., 1992; Dickinson et al., 1997). Die Entwicklung von *in-silico* Vorhersageprogrammen für S/MAR-Elemente beruht auf der Suche nach solchen Sequenz-elementen (Benham, 1996). Das Genomatix-Programm „ElDorado“ nutzt das Programm „SMARTest“ zur Vorhersage potentieller S/MAR-Elemente (Frisch et al., 2002; Liebich et al., 2002; Evans et al., 2007). Dies geschieht anhand der Gewichtungsmatrizen von 97 experimentell bestätigten S/MAR-Elementen. Zur S/MAR-Vorhersage wird von „SMARTest“ eine mindestens 300 bp lange Kandidaten-Sequenz benötigt. Ein zweites online-basiertes S/MAR-Analyseprogramm „MARFinder“ (Singh et al., 1997) wurde zum Vergleich der „SMARTest“-Ergebnisse eingesetzt.

Es wurden die MARFinder Standardeinstellungen, unter Anwendung von sechs potentiellen Erkennungsregeln[51], verwendet. Die „SMARTest" Analyse ergab für den *DDX3Y* Genlocus, innerhalb des untersuchten 50 kb Fensters, acht potentielle S/MAR-Elemente. Zwei S/MARs (Nr. 1 und 2) sind in ~ 15,23 kb und 4,2 kb Entfernung stromaufwärts zum *DDX3Y* ATG lokalisiert (Abb. 4-53; Tab. 4-26). Eine dritte potentielle S/MAR-Sequenz liegt intergenisch in *DDX3Y* Intron 4. Die vierte potentielle S/MAR-Sequenz liegt in der 3´UTR von *DDX3Y* Exon17. In 8,78 kb bis 15,24 kb Entfernung stromabwärts vom *DDX3Y* Exon17 3´Ende (Pos. 15.032.390) sind vier weitere potentielle S/MAR-Elemente vorhergesagt (Tab. 4-26). Die potentiellen S/MARs Nr. 6 bis 8 sind bereits innerhalb der nachfolgenden Y-chromosomalen Transkriptionseinheit lokalisiert. Hierbei handelt es sich um das unprozessierte, transkribierte Pseudogen *CASKP1* (Pos. 15.042.075 - 15.060.090, reverser Strang).

Tab. 4-26: Regionen im *DDX3Y* Locus mit potentiellen S/MAR-Elementen.

Kartierung von potentiellen S/MAR-Elementen in 50 kb der Human Y-Chromosom-Sequenz NC_000024.9 (Position 14.998.955 - 15.048.954) mit dem Genomatix-Programm „SMARTest". Die *DDX3Y* Exon-Intronstruktur erstreckt sich von Pos. 15.015.753 (T-TSS-II; relative Pos. in 50 kb Seq. = 16.798) bis 15.032.390 (3´Ende Exon17; relative Pos. in 50 kb Seq. = 33.435). Das Pseudogen *CASKP1* beginnt in Position 15.042.075 (relative Pos. in 50 kb Seq. = 43.120).

Nummer	Anfang	Ende	rel. Position in 50 kb Seq.	Länge in bp	Beschreibung
1	15.001.621	15.001.985	2.666	365	5´
2	15.012.006	15.012.440	13.051	435	5´
3	15.024.146	15.024.470	25.191	325	*DDX3Y* Intron 4
4	15.031.276	15.031.655	32.321	380	*DDX3Y* Exon 17
5	15.041.161	15.041.965	42.206	805	3´
6	15.043.606	15.044.085	44.651	480	*CASKP1* Intron7
7	15.045.596	15.046.050	46.741	455	*CASKP1* Intron7
8	15.047.176	15.047.625	48.221	450	*CASKP1* Exon7 + Intron6

Mit dem zweiten Programm „MARFinder" konnten drei potentielle S/MAR-Sequenzen, mit Peakstellen in den relativen Positionen 2.500 (Pos. 15.001.455), 12.800 (Pos. 15.011.755) und 42.700 (Pos. 15.041.655), kartiert werden (Abb. 4-55). Nur die potentiellen S/MARs Nr. 5 (S/MARTest) und S/MAR-42.700 (MARFinder) ko-lokalisieren und werden von beiden Programmen vorhergesagt (Tab. 4-26, grau unterlegt). Dieses pot. S/MAR-Element ist direkt 5´ der *CASKP1* Transkriptionseinheit zwischen *DDX3Y* und *CASKP1* lokalisiert. Die S/MARs Nr. 1 und 2 überlappen geringfügig mit den S/MARs-2.500 und -12.800, ko-lokalisieren aber nicht über die gesamte Länge (Abb. 4-55).

[51] MARFinder-Regeln: „Origin of Replication", TG-Gehalt, AT-Gehalt, gebogene und geknickte DNA-Konformationen, Topoisomerase II-Erkennungssequenz.

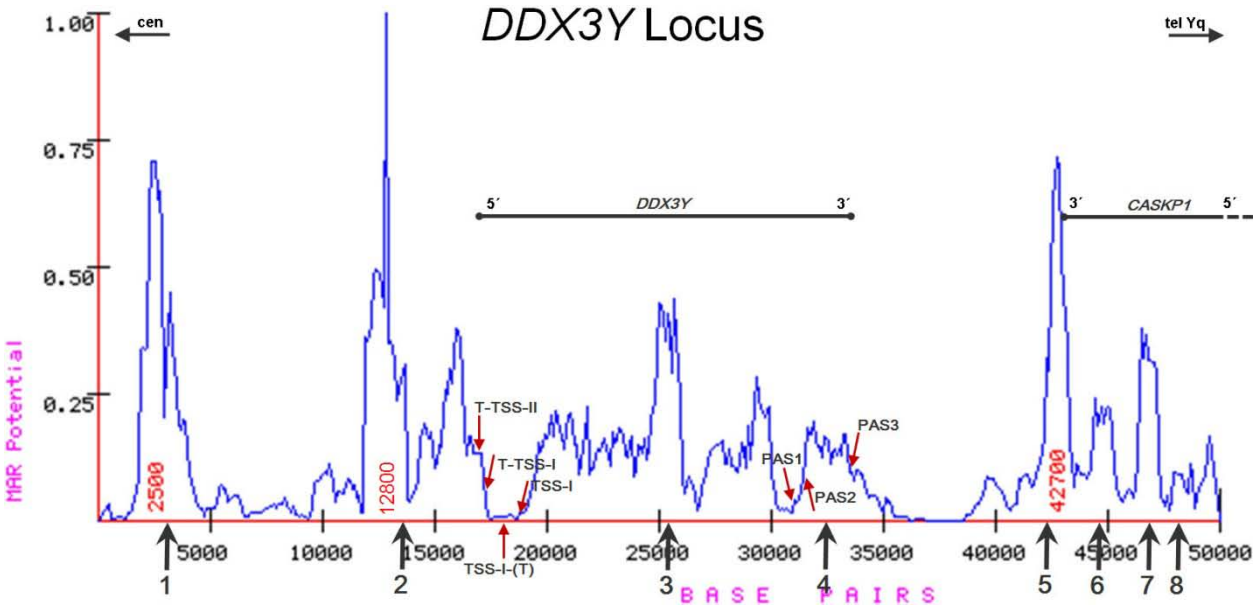


Abb. 4-55: Karte der MAR-Vorhersage mit MARFinder für den *DDX3Y* Locus.

Kartierung von potentiellen S/MAR-Elementen in 50 kb der Human Y-Chromosom-Sequenz NC_000024.9 (Position 14.998.955 - 15.048.954) durch MARFinder-Programm. Drei potentielle S/MARs mit Peakstellen in 2.500 (Pos. 15.001.455), 12.800 (Pos. 15.011.755) und 42.700 (Pos. 15.041.655) wurden identifiziert. Signifikante S/MAR-Sequenzen liegen oberhalb des Grenzwertes von 0,6. Die Pfeile 1 bis 8 geben die Positionen der pot. S/MAR-Elemente an, die mit S/MARTest kartiert wurden. Die roten Pfeile markieren die *DDX3Y* TSS-Positionen, sowie die PAS-Positionen. Der linke horizontale Balken gibt den *DDX3Y* Genbereich an. Der rechte Balken gibt den *CASKP1* Pseudogenbereich an. *CASKP1* expandiert weiter 3´ (durch gestrichelten Balken angezeigt). Die horizontalen Pfeile geben die Orientierung im Chromosom an.

Für *DDX3X* konnte mit „SMARTest“ nur ein potentielles S/MAR-Element in der 3´UTR-Sequenz von Exon17 an Position 41.208.496 (370 bp; relative Pos. 32.656) identifiziert werden (Abb. 4-56). Die Position ist homolog zur potentiellen S/MAR-Sequenz Nr. 4 im *DDX3Y* Exon17. Vergleichbar zum potentiellen *DDX3Y* S/MAR-Element in Exon17, weist das DDX3X S/MAR-Element nur ein sehr geringes S/MAR-Potential auf, welches deutlich unterhalb des Signifikanz-Grenzwertes von 0,6 liegt (Abb. 4-56). Mit „MARFinder“ konnten zwei potentielle S/MARs in den Positionen 41.198.240 (relative Pos. 22.400) und 41.219.240 (relative Pos. 43.400) kartiert werden (Abb. 4-56). S/MAR-22400 beginnt in *DDX3X* Intron2 und endet in Intron3. Das S/MAR-Element 43.400 ist in einem *Alu*-Sequenzblock zwischen *DDX3X* Exon17b und 17c lokalisiert. Die insgesamt drei potentiellen S/MAR-Sequenzen im *DDX3X* Locus, die mit zwei Programmen ermittelt wurden, ko-lokalisieren nicht.

In den potentiellen *DDX3Y* und *DDX3X* S/MAR-Sequenzen konnten jeweils mehrere potentielle MARBPs mit MatInspector kartiert werden. Dabei wurden unterschiedliche Kombinationen der MARBPs SATB1, ARID3A (AT-rich interactive domain-containing protein 3A = BRIGHT), ARID5A (= MRF1), HMG-I/Y (= HMGA1) und CDP (CCAAT displacement protein = CUX1) identifiziert. Ein gewebespezifisches MARBP ist z.B. der Faktor SATB1, der auch in geringen

Mengen in Hodengewebe zu finden ist, und spezifisch an doppelsträngige BURs bindet (Wang et al., 2010; Cai et al., 2003; Dickinson et al., 1992).

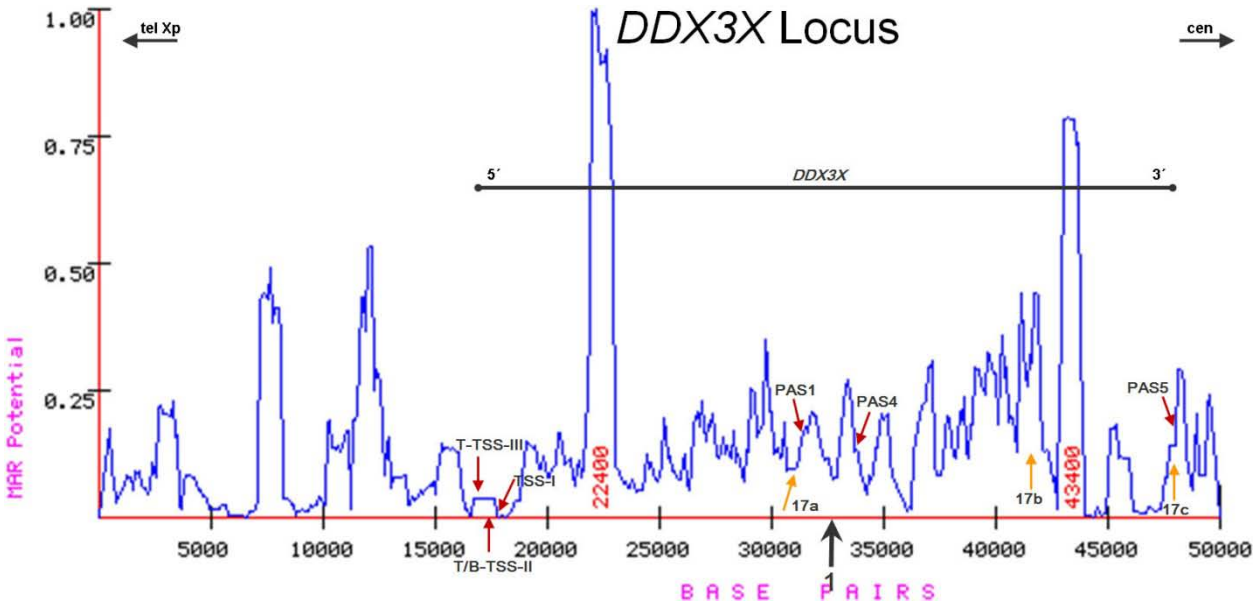


Abb. 4-56: Karte der MAR-Vorhersage mit MARFinder für den *DDX3X* Locus.
Vorhersage in 50 kb Human X-Chromosom-Sequenz NC_000023.10 (Position 41.175.840 - 41.225.839) durch MARFinder-Programm. Zwei potentielle S/MARs mit Peakstellen in 22.400 (Pos. 41.198.240), und 43.400 (Pos. 41.219.240). Signifikante S/MAR-Sequenzen liegen oberhalb des Grenzwertes von 0,6. Der Pfeil mit der Markierung 1 gibt die Position des pot. S/MAR-Elementes an, das mit S/MARTest kartiert wurde. Die roten Pfeile markieren die *DDX3X* TSS-Positionen, sowie die PAS-Positionen. Die gelben Pfeile markieren die Positionen der gesplicten, hodenspezifischen Exon17 Transkriptvariante. Der horizontale Balken gibt den *DDX3X* Genbereich an. Die horizontalen Pfeile geben die Orientierung im Chromosom an.

In Kapitel 4.6.3.1 wurde eine konservierte SATB1-Bindestelle in der X-Y konservierten „KRiH-1" Sequenz festgestellt. Die „KRiH-1"-Sequenz wurde von „SMARTest", weder in der *DDX3Y* (15.014.561 - 15.014.762; relative Pos. in 50 kb Seq. = 15.606 - 15.807), noch in der *DDX3X* Sequenz (41.191.335 - 41.191.559; relative Pos. in 50 kb Seq. = 15.495 - 15.719), als potentielles S/MAR-Element identifiziert. Dieser Unterschied in der Identifizierung, zwischen den Programmen MatInspector und S/MARTest, entsteht wahrscheinlich durch die unterschiedlichen Suchalgorithmen für eine solche Sequenz. Eine TFBS ist ~ 15 bp lang, ein potentielles S/MAR mindestens 300 bp. In Abb. 4-55 und Abb. 4-56 ist zu sehen, daß die *DDX3Y* und *DDX3X* KRiH-1 Sequenzen unterhalb des MARFinder Grenzwertes von 0,6 liegen und somit auch mit dem zweiten Programm keine S/MAR-Sequenz im Bereich von KRiH-1 vorhergesagt wird.

Für die 2.545 bp großen genomischen *DDX3Y* und *DDX3X* Bereich, die in Kapitel 4.6. untersucht wurden, konnten mehrere potentielle CTCF/BORIS TFBS festgestellt werden (Tab. A-16, A17). CTCF und das paraloge BORIS sind, wie auch die MARBPs, ebenfalls an der Chromatinorganisation beteiligt und gelten als Faktoren mit Insulator-Funktion (Klenova et al.,

2002; Dunn et al., 2003; Williams & Flavell, 2008). Beide Zinkfingerfaktoren erkennen dieselben *cis*-Elemente auf der DNA (Klenova et al., 2002). Das Insulatorprotein BORIS wird in Mensch und Maus hodenspezifisch exprimiert und es konnte auch gezeigt werden, daß CTCF/BORIS an MARs binden können (Dunn & Davie, 2003; Hore et al., 2008; Ohlsson et al., 2010). Die drei potentiellen TFBS in der *DDX3Y* und acht in der *DDX3X* Promoterregion erreichen nicht den empfohlenen Matrix-Wert von 0,85 (Tab. A-16, A17), für eine funktionelle TF-Bindung und es kann, zumindest für den 2,5 kb Bereich eine Insulatorfunktion bei der Chromatinorganisation und Genkontrolle ausgeschlossen werden.

Interessanterweise konnte am 3´Ende des S/MAR-Elementes Nr. 8 im *DDX3Y* Genlocus eine 201 bp lange Sequenz (Pos. 15.047.537 - 15.047.737) festgestellt werden, in der 28 potentielle TFBS für einen GAGA-Faktor vorhanden sind (Tab. A-28). GAGA-Faktoren können durch ihren Einfluß auf die Chromatinarchitektur als Insulator-Elemente wirken (West et al., 2002). Für die Position dieses potentiellen S/MAR-Elementes konnte in Abb. 4-55 nur ein sehr geringes Potential zur funktionellen S/MAR-Ausbildung festgestellt werden. Da die Position von S/MAR Nr. 8 zudem bereits innerhalb der *CASKP1* Transkriptionseinheit lokalisiert ist, könnte es eher als reiner Insulator und nicht als S/MAR-Element funktionieren. Diese Region könnte also zwei Chromatindomänen funktionell voneinander trennen.

Alu-Elemente:

Als zweites Sequenzelement wurden repetitive Sequenzblöcke der *Alu*-Familie verglichen, da *Alu*-Elemente, mit über 1 Mio. Kopien pro Humangenom, die zahlenmäßig am häufigsten vorkommenden transponierbaren Elementen (TE) darstellen (Cordaux & Batzer, 2009). *Alu*-Elemente gehören zu den primatenspezifischen SINE („short interspersed nuclear elements“) Retrotransposons. Sie sind typischerweise zwischen 100 und 300 bp lang und relativ frei über das gesamte Humangenom verteilt (Houck et al., 1979; Rowold & Herrera, 2000; Batzer & Deininger, 2002). *Alu*-Elemente sind nicht autonom mobil, sondern leihen die „Reverse Transkriptase“ der LINEs („long interspersed nuclear elements“) für ihre Mobilität aus (Batzer & Deininger, 2002). Als Herkunft der primatenspezifischen *Alu*-Elemente werden 7SL-RNAs vermutet, die auch als Ursprung der Nagetier-spezifischen B1 SINE-Familie gelten (Kriegs et al., 2007).

Alu-Elemente tragen noch heute zur Veränderung des Genomes bei, indem neue *Alu*-Sequenzen eingebaut werden und homologe Rekombinationen zwischen entfernt liegenden *Alu*-Sequenzen zu Duplikationen, Deletionen und Translokationen führen können (Batzer & Deininger, 2002). Durch *Alu*-Repeats können zudem neue SSRs („Simple Sequence Repeats“) entstehen. In der Mitte der *Alu*-Elemente kommt eine A-reiche Region (mit der Sequenz A_5TACA_6) vor und am Ende des *Alu*-Elementes integriert der poly(dA)-Schwanz, der einen

homopolymeren A-Repeat von bis zu 100 Nukleotiden erzeugen kann (Batzer & Deininger, 2002).

Eine weitere Eigenschaft der *Alu*-Elemente ist der relativ hohe (C,G)-Gehalt. (C,G)-reiche Sequenzen wie *Alu*-Elemente sind im Gegensatz zu CpG-Islands normalerweise dauerhaft methyliert (Caiafa & Zampieri, 2005) und dienen als *de novo* Methylierungscenter, von denen *de novo* Methylierungen auf umliegende genregulative Sequenzen übergreifen können (Caiafa & Zampieri, 2005). Dies ist eine entgegengesetzte Funktion zu den oben beschriebenen S/MAR-Elementen, die *de novo* Methylierungen verhindern können.

Der Methylierungsstatus von *Alu*-Elementen kann zudem gewebespezifisch und temporal unterschiedlich sein (Batzer & Deininger, 2002). Besonders die Methylierungsmuster zwischen hypermethylierten *Alu*-Sequenzen im Soma und hypomethylierten in der männlicher Keimbahn weisen große Unterschiede auf (Hellmann-Blumberg et al., 1993; Rubin et al., 1994; Chesnokov & Schmid, 1995). Kvikstad & Makova (2010) konnten zudem eine bevorzugte Integration von *Alu*-Elementen in die männliche Keimbahn zeigen.

In der analysierten 50 kb Sequenz des *DDX3Y* Genlocus konnten 20 *Alu*-Elemente (9 auf dem +Strang, 11 auf dem -Strang) festgestellt werden (Abb. 4-53). Stromaufwärts des *DDX3Y* ATGs konnten zwei Blöcke mit sieben bzw. vier *Alu*-Elementen kartiert werden. Im ersten distalen Block beträgt der mittlere Abstand der *Alu*-Sequenzen 1380 bp. Im zweiten Block zwischen S/MAR Nr. 2 und *MSY2* Repeat beträgt der mittlere Abstand 480 bp. Weitere *Alu*-Elemente befinden sich in *DDX3Y* Intron2, 3, 6 und 16. Zwei *Alu*-Sequenzen befinden sich 8,54 kb stromabwärts des *DDX3Y* Exon17 in unmittelbare Nähe der S/MAR-Sequenz Nr. 5. Zwischen S/MAR Nr. 6 und 7 und 3´ distal zu S/MAR Nr. 8 ist jeweils eine *Alu*-Sequenz vorhanden.

In der 50 kb *DDX3X* Genlocus-Sequenz konnten 51 *Alu*-Elemente (21 auf dem +Strang, 30 auf dem -Strang) kartiert werden (Abb. 4-53). Stromaufwärts des *DDX3X* ATGs konnten in einer Entfernung von 17,33 kb und 10,4 kb zwei Blöcke mit zehn und zwölf *Alu*-Elementen identifiziert werden. Der erste Block umfaßt eine Region von 5,13 kb, der zweite von 6,77 kb. Der mittlere Abstand der *Alu*-Sequenzen beträgt 491 bp in Block 1 und 556 bp in Block 2. Intergenische *Alu*-Elemente sind in *DDX3X* Intron1, 2, 3 und 10 lokalisiert.

Zwischen *DDX3X* Exon17 und dem Exon17b ist ein 4,5 kb langer Block mit neun *Alu*-Elementen und einem mittleren Abstand von 482 bp kartierbar. Zwischen Exon17b und 17c ist nochmals ein 5,64 kb langer Block aus elf *Alu*-Sequenzen mit mittlerem Abstand der Elemente von 492 bp lokalisiert. Stromabwärts von *DDX3X* Exon17c sind vier weitere *Alu*-Sequenzen mit mittleren Abständen von 367 bp vorhanden (Abb. 4-53).

Die *Alu*-Sequenzen in den *DDX3Y* und *DDX3X* Loci sollten nicht homologen Ursprunges sein, da die Trennung der Gonosomen vor der Entstehung der primatenspezifischen *Alu*-Sequenzen

stattgefunden hat. Möglicherweise bestand aber die Möglichkeit zu partiellen Rekombinationen über repetitive Sequenzen zwischen den Gonosomen, bzw. zu intrachromosomalen Rekombinationen.

Der Vergleich der S/MAR- und *Alu*-Sequenzelemente in den *DDX3Y* und *DDX3X* Genloci zeigt, trotz der sehr hohen Konserviertheit der *DDX3Y* und *DDX3X* Exon- und Proteinsequenzen, also deutliche Unterschiede in der Evolution beider Genombereiche, die über die Promoterdomänen hinausgehen. Auf Grund der höheren Anzahl an potentiellen (A,T)-reichen S/MARs kann für *DDX3Y* ein Modell der Expressionskontrolle über gewebespezifisch ausgebildete Chromatin-schleifen vorgeschlagen werden. Zudem ist die gesamte *DDX3Y* Exon-Intron-Struktur zwischen zwei potentiellen S/MAR-Elementen (MARFinder 12.800, 42.700) lokalisiert (Abb. 4-55).
Für *DDX3X* wurden drei potentielle S/MAR-Sequenzen kartiert. Im Unterschied zu *DDX3Y* konnten für den *DDX3X* Genlocus mehr als doppelt so viele (C,G)-reiche *Alu*-Elemente kartiert werden. Eine Eigenschaft von (G+C)-reichen DNA-Sequenzen ist der Ausschluß aus Regionen der Chromatin-loop Organisation, den (A+T)-reichen S/MARs. Dies hat zur Folge, daß vor allem „Housekeeping-Gene“, die häufig über CpG-Islands reguliert werden, nicht in den Transkriptions-inaktiven Binderegionen der S/MARs zu liegen kommen (Beggs & Migeon, 1989; Caiafa & Zampieri, 2005).

4.8.3. *DDX3X* Intron1 Gen

Ein weiterer Unterschied der beiden Genloci konnte im *DDX3X* Genlocus festgestellt werden (Abb. 4-53). Nur hier wurde ein 2129 bp langes, intronloses Gen[52] im *DDX3X* Intron1 mit einem offenen Leserahmen von 185 AS identifiziert. Eine entsprechende Sequenz im *DDX3Y* Intron1 konnte nicht gefunden werden. Mit dem Primerpaar P2033/2034 (RT-PCR Amplifikat = 373 bp) konnte festgestellt werden, daß dieses Transkript weder hoden- noch geschlechtsspezifisch transkribiert wird (Abb. 4-57). Alle bisherigen RT-PCRs für *DDX3X* zeigen, daß dieses Transkript eigenständig ist und nicht in *DDX3X* Transkripten, als alternatives Exon, integriert wird (P1544/2034; Abb. 4-57). P1544 ist in *DDX3X* Exon1 lokalisiert. Um eine Kontamination der cDNA mit DNA auszuschließen, wurden die RNA vor der cDNA-Umschreibung mit denselben Primerpaaren getestet. Es konnten keine PCR-Produkte festgestellt werden (nicht gezeigt). Innerhalb der 3´UTR Sequenz konnte ein *Alu*-Element kartiert werden (Abb. 4-53). Um Kreuzreaktionen mit anderen *Alu*-Elementen zu vermeiden, wurden die Primer P2033 und P2034 außerhalb dieser *Alu*-Sequenz positioniert.

---

[52] GenBank Acc.No. „XM_001129278.1“; Pos. 41.193.863 - 41.195.991; 2.129 bp Länge mit 178 bp 5´UTR, 555 bp ORF und 1.396 bp 3´UTR.

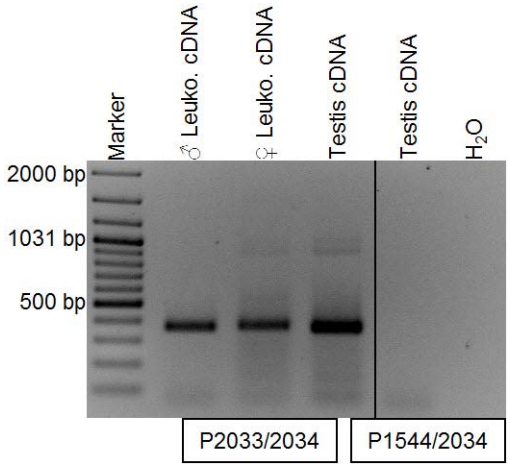


Abb. 4-57: RT-PCR Analyse *DDX3X* Intron1 Gen.
Errechnete PCR-Produktgröße für P2033/2034 = 373 bp, für P1544/2034 = 1400 bp. P2033 (BAC-Pos. 31.738) und P2034 (BAC-Pos. 32.110) liegen außerhalb der *Alu*-Repeats. P2033 ist im potentiellen ORF, P2034 in der potentiellen 3´UTR lokalisiert. P1544 (BAC-Pos. 30.711 ist im ubiquitären *DDX3X* Exon1 vor dem ATG lokalisiert. 35 PCR Zyklen. Als DNA-Marker wurde eine 100bp DNA-Leiter verwendet.

Die von Vinckenbosch et al. (2006) und Shiao et al. (2007) beschriebenen Eigenschaften eines Retrogens[53] treffen auf das *DDX3X* Intron1 Gen zu.

Evolutionär scheint dieses *DDX3X* Intron1 Retrogen bereits vor ~ 100 Mio. Jahren, vor der Trennung der Euarchontoglires und der Laurasiatheria (Hedges & Kumar, 2003; Kriegs et al., 2006; Janečka et al., 2007), in die X-chromosomale Linie integriert worden zu sein. Der Leserahmen ist in Schimpanse, *Macaca*, *Callithrix*, Rind und Maus wiederzufinden. Das ATG in adäquatem Kontext ist in allen sechs Spezies konserviert. Die ORF-Länge und vor allem die ORF-Sequenzen divergieren allerdings auf Grund von indels sehr stark. In Schimpanse beträgt die ORF-Länge 84 AS, in *Macaca* und *Callithrix* 17 AS, beim Rind 141 AS und bei der Maus 148 AS. Die potentiellen 5´- und 3´UTR-Sequenzen zeigen, vor allem zwischen den Primaten, einen hohen Grad an Homologie.
Während der Primatenevolution integrierte, vor mind. 35 - 43 Mio. Jahren, vor der Trennung der Catarrhini und Platyrrhini (Chatterjee et al., 2009), dann zusätzlich ein *Alu*-Element in die 3´UTR-Sequenz dieses Retrogens. Das primatenspezifische *Alu*-Element ist in Rind und Maus entsprechend nicht zu finden. Eine konservierte Funktion des potentiellen Proteins ist auf Grund der vielen Veränderungen unwahrscheinlich. Eine Funktion als regulative RNA, für z.B. die Kontrolle der *DDX3X* Transkription oder Translation, kann nicht ausgeschlossen werden.

[53] Retrogeneigenschaften: Intronlos, Integration in Intronsequenz eines anderen Genes und Anwesenheit einer genomischen Poly-A-Sequenz im Anschluß an die 3´UTR Sequenz.

4.9. Funktionelle Analysen zur Human *DDX3X* Promoterregion

Nachdem in Kapitel 4.2. gezeigt werden konnte, daß die *DDX3X* 5´Region über mehrere Transkriptionsinitiationsstellen verfügt und somit Bestandteil einer komplexen Promoterdomäne ist, soll in diesem Kapitel die potentielle *DDX3X* Promoterregion auf funktionelle Subdomänen, mit Hilfe des Luziferase-Reportergensystems, untersucht werden. Dazu wurde das „Dual-Luciferase Reporter Assay" System (DLR) mit pGL4-Vektoren gewählt. Zur Analyse des *DDX3X* Promoter-Expressionsprofils wurden zwei Testserien von potentiellen *DDX3X* Promoter-Subfragmenten in diesen Vektor kloniert. Die erste Serie bestand aus zehn *DDX3X* Promoter-Reportergenkonstrukten mit Insertlängen von 0,41 bis 2,44 kb. Damit sollten die potentielle 5´Grenze der funktionellen *DDX3X* Promoterregion, sowie Regionen hoher und niedriger Promoteraktivität definiert werden (Abb. 4-58). Die zweite Serie bestand aus 13 *Luc*-Reportergenkonstrukten, mit denen einzelne Abschnitte, der kartierten *DDX3X* Promoterregion, gezielt auf spezifische Aktivator- und Repressoreffekte analysiert wurden (Abb. 4-60). Zur Auswahl der optimalen *DDX3X* Subfragmente in beiden Testserien, wurden die Daten aus der *DDX3X* Core-Promoteranalyse (Kapitel 4.2), sowie der *in-silico DDX3X* Promoterkartierung (Kapitel 4.6) genutzt.

Als Negativkontrolle, wie auch als Promoter-Klonierungsvektor beider Testserien, diente der endogen promoterlose Vektor pGL4.10 ([*luc2*]; GenBank Acc.No. AY738222.1). Dieser enthält nur das synthetische Firefly-Luziferasegen [*luc2*]. Der Vektor pGL4.13 ([*luc2*/SV40]; GenBank Acc.No. AY738225), mit zusätzlich endogener SV40 („Simian-Virus 40") Promoter- und Enhancerregion, diente als Positivkontrolle, sowie als zweiter Klonierungsvektor für die Kartierung potentiell Expressions-verstärkender oder -reprimierender *DDX3X* Promoter-abschnitte. Um mögliche Fehler durch unterschiedliche Transfektionseffizienzen in den verschiedenen Experimentserien zu minimieren, wurde in allen Experimenten die Reportergen-aktivität der Firefly-Luziferase mit der co-transfizierten *Renilla*-Luziferase des Co-Reporter-vektors pGL4.74 ([*hRluc*/TK]; GenBank Acc.No. AY738230) normalisiert. Die *Renilla*-Luziferaseexpression wird von einem HSV-TK Promoter (Herpes simplex Virus Thymidin-Kinase) kontrolliert.

4.9.1 Funktionelle Kartierung der *DDX3X* Promoterdomäne

Die Analyse der *DDX3X* Promoterregion mit sequentiell 5´ verlängerten Abschnitten der *DDX3X* Promotersequenz ist eine Fortführung und Erweiterung der Ergebnisse von Hanstein (2004). Die zehn *DDX3X* Promotersequenzen, die in den Reportervektor pGL4.10 kloniert wurden, sind

graphisch in Abb. 4-58 und tabellarisch in Tab. 4-27 dargestellt. Die *DDX3X* Sequenzen wurden mittels zweier Restriktionsenzyme stromaufwärts des *luc2*-Reportergens in die „Multiple Cloning Site“ (MCS) des pGL4.10 Vektors integriert. In die Forwardprimer wurde immer eine Erkennungssequenz für *SacI* („GAGCTC“) an das 5´ Primerende, an den Reverseprimer *NheI* („GCTAGC“) als Erkennungssequenz angehängt. Alle Promoterkonstrukte wurden nach erfolgreicher Klonierung komplett sequenziert, um korrekte *DDX3X* Promotersequenzen zu gewährleisten.

Tab. 4-27: Liste der *DDX3X* Promoterkonstrukte.

Die *DDX3X* Promotersequenzen sind jeweils schrittweise zwischen ~ 140 und 320 bp 5´ verlängert. Die Konstrukte enden am 3´Ende jeweils fix im *DDX3X* Exon1 (BAC-Position 30.764). Die zehn *DDX3X* Promoterkonstrukte sind von distal nach proximal mit römischen Ziffern durchnummeriert.

Nummerierung	Konstruktbezeichnung	Position im BAC RP13-13A3	Länge der klonierten Fragmente
I	pGL4.10-2,44 kb	28.323 - 30.764	2.442 bp
II	pGL4.10-2,13 kb	28.638 - 30.764	2.127 bp
III	pGL4.10-1,99 kb	28.779 - 30.764	1.986 bp
IV	pGL4.10-1,85 kb	28.915 - 30.764	1.850 bp
V	pGL4.10-1,57 kb	29.198 - 30.764	1.567 bp
VI	pGL4.10-1,31 kb	29.460 - 30.764	1.305 bp
VII	pGL4.10-1,11 kb	29.658 - 30.764	1.107 bp
VIII	pGL4.10-0,92 kb	29.846 - 30.764	919 bp
IX	pGL4.10-0,63 kb	30.138 - 30.764	627 bp
X	pGL4.10-0,41 kb	30.360 - 30.764	405 bp

Die experimentelle Auswertung der pGL4.10-Promoterkonstrukte in Abb. 4-59 zeigt, daß die unterschiedlich langen *DDX3X* Promotersequenzen unterschiedlich starke Luziferase-aktivitäten erzeugen. Die absoluten und relativen Meßwerte sind in Tabelle 4-28 eingetragen. Die schwächste rel. Luziferaseaktivität mit 7,9% wurde für die kürzeste *DDX3X* Promoter-sequenz (Nr. X) festgestellt. Mit 89,7% rel. Luziferaseaktivität wurde für das Konstrukt VI (1,31 kb) die stärkste Aktivität gemessen. Anhand Abb. 4-58 kann abgelesen werden, welche potentiellen Promoterbestandteile in diesen *DDX3X* Promotersequenzen vorhanden sind. So sind in Konstrukt X nur das kurze, potentielle CpG-Island „CpG-P2“ (Kapitel 4.6.2) und der ubiquitäre Transkriptionsstart *DDX3X* TSS-I enthalten. In dem Konstrukt VI sind alle drei *DDX3X* TSS-Positionen, fast die komplette CpG-Island Region „CpG-IS“, die *MSY2-X* Sequenz, die drei Regionen mit repetitiven TF-Bindestellen, sowie alle kartierten auffälligen Sequenz-trakte, außer den distalen W-Trakten, vorhanden. Mit Konstrukt Nr. VIII (0,92 kb) konnte eine rel. Luziferaseaktivität von 77,3% in Bezug zum pGL4.13-Vektor festgestellt werden. In den verschiedenen Replikaten fiel auf, daß mit Konstrukt VII (1,1 kb) eine geringere Luziferase-aktivität erzielt wird, als mit den beiden flankierenden Konstrukten VI und VIII.

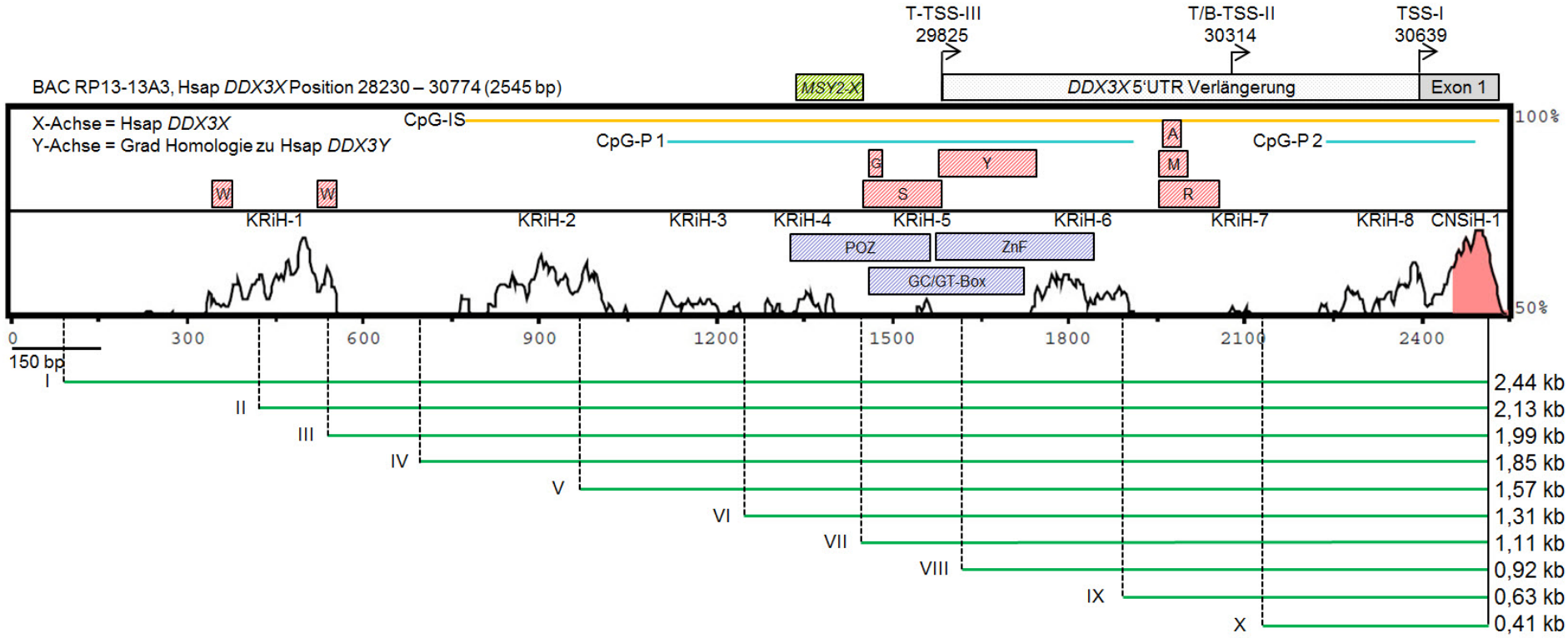


Abb. 4-58: Graphische Darstellung der *DDX3X* Promoterkonstrukte.

Die Graphik basiert auf den Ergebnissen der *in-silico* Promoteranalyse in Kapitel 4.6. Die zehn *DDX3X* Konstrukte sind durch römische Ziffern gekennzeichnet, die in Tabelle 4-27 wiederzufinden sind. Positionsangaben beziehen sich auf den Human BAC-Klon RP13-13A3. *MSY2-X* = BAC-Pos. 29.570 - 29.676; W1 = BAC-Pos. 28.574 - 28.603; W2 = BAC-Pos. 28.741 - 28.780; G = BAC-Pos. 29.681 - 29.692; S = BAC-Pos. 29.675 - 29.821; Y = BAC-Pos. 29.816 - 29.978; A = BAC-Pos. 30.189 - 30.213; M = BAC-Pos. 30.186 - 30225; R = BAC-Pos. 30.186 - 30.287; CpG-IS = BAC-Pos. 29.011 - 30.774; CpG-P1 = BAC-Pos. 29.349 - 30.144; CpG-P2 = BAC-Pos. 30.467 - 30.717; POZ-TF = BAC-Pos. 29.552 - 29.812; ZnF-TF = BAC-Pos. 29.813 - 30.090; SP1-TF = BAC-Pos. 29.681 - 29.932; KRiH-1 = BAC-Pos. 28.559 - 28.783; KRiH-2 = BAC-Pos. 28.994 - 29.257; KRiH-3 = BAC-Pos. 29.333 - 29.432; KRiH-4 = BAC-Pos. 29.515 - 29.629; KRiH-5 = BAC-Pos. 29.766 - 29.796; KRiH-6 = BAC-Pos. 29.951 - 30.142; KRiH-7 = BAC-Pos. 30.302 - 30.352; KRiH-8 = BAC-Pos. 30.452 - 30.670; CNSiH-1 = BAC-Pos. 30.671 - 30.774.

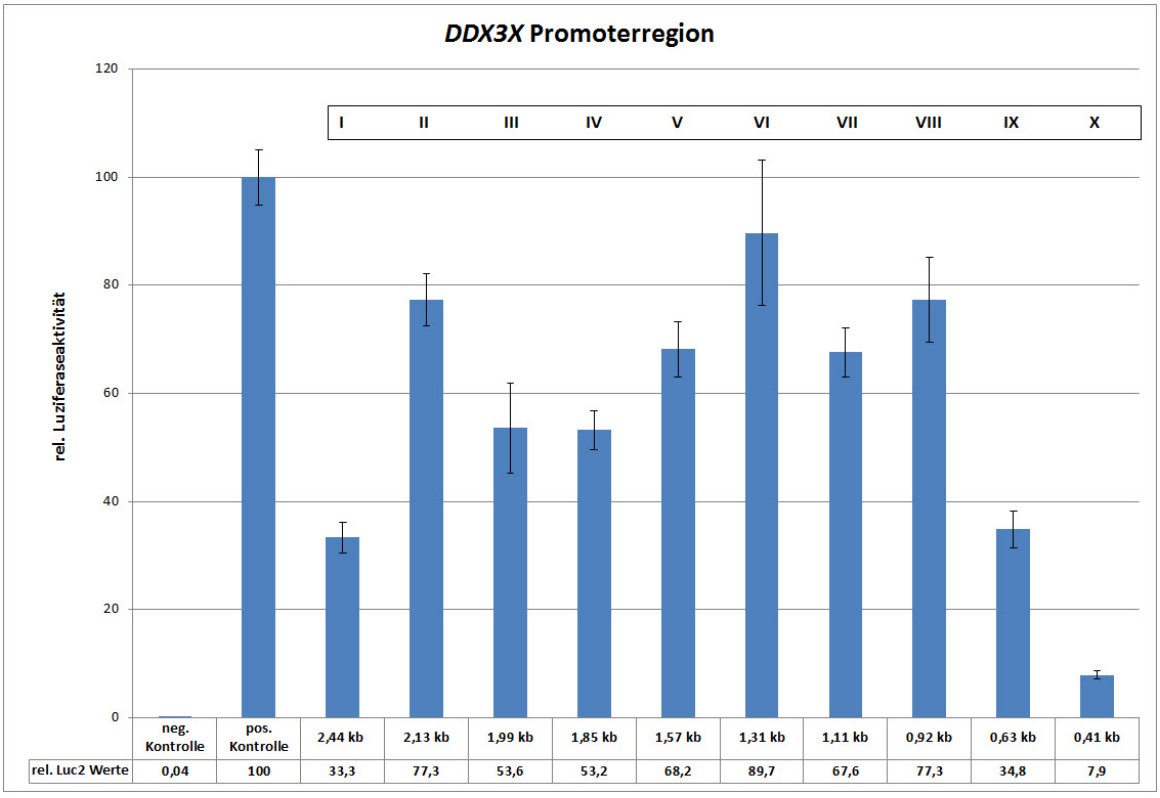


Abb. 4-59: Auswertung des Human *DDX3X* Promoterregion DLR-Assays.
Transiente Transfektion in die Tumorzellinie EP2102. Experimenteller Vektor ist der endogen promoterlose Vektor pGL4.10. Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3X* Promoterkonstrukte I bis X dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-28 aufgeführt.

Distal des stärksten Konstruktes Nr. VI ist eine kontinuierliche Abnahme der Luziferaseaktivität bis zum Konstrukt III (1,99 kb) feststellbar. Für das Konstrukt II (2,1 kb) konnte ein erneuter Anstieg der rel. Luziferaseaktivität auf 77,3% festgestellt werden, die vergleichbar mit der Luziferaseaktivität von Konstrukt VIII ist. Konstrukt II enthält bis auf die distale W1-Sequenz alle kartierten potentiellen Promoterelemente. Für das längste *DDX3X* Promoterkonstrukt I (2,44 kb) konnte nur noch eine rel. Aktivität von 33,3% gemessen werden. Längere *DDX3X* Promotersequenzen konnten auf Grund von Rekombinationsereignissen innerhalb der *DDX3X* Promotersequenz nicht stabil kloniert werden.

Ein inhärentes Problem der durchgeführten DLR-Assays sind die schwankenden *Renilla*-Meßwerte. Die absoluten *Renilla*-Werte weichen bis zu Faktor 3 von dem *Renilla*-Meßwert der Positivkontrolle ab (Tab. 4-28). Von *Renilla* Co-Reportervektoren mit stärkeren Promoter-sequenzen wie SV40 oder CMV ist bekannt, daß sie in Konkurrenz zum experimentellen Vektor treten können und beide transfizierten Vektoren um verfügbare *trans*-Faktoren konkurrieren. Diese „Promotor-Crosstalk Interferenz" wird zu Gunsten des stärkeren Promoters entschieden, der die allgemeinen wie auch spezifischen Transkriptionsfaktoren aus dem System abfangen

kann („Squelching Effect"; Behre et al., 1999; Matis et al., 2001; Zhuang et al., 2001; Hampf & Gossen, 2007). Aus diesem Grund wurde hier bewußt der pGL4.74 Vektor mit HSV-TK Promoter gewählt, da dieser in der Regel niedrige bis moderate, dafür aber relativ uniforme *Renilla*-Luziferase Expressionen in transfizierten Säugerzellinien liefert.

Tab. 4-28: Absolute und relative Meßwerte der *DDX3X* Promoterregion DLR-Analyse.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug dazu dargestellt. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Versuchszeitpunkte.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
I) pGL4.10-2,44 kb	1.161.402,7	75.031,3	154.999,9	**33,3**	2,8
II) pGL4.10-2,13 kb	4.853.859	134.664	360.433,3	**77,3**	4,9
III) pGL4.10-1,99 kb	1.674.323	68.107	249.879,9	**53,6**	8,3
IV) pGL4.10-1,85 kb	4.896.542,3	197.853	248.199,3	**53,2**	3,6
V) pGL4.10-1,57 kb	5.089.901,5	161.130,8	317.808,7	**68,2**	5,1
VI) pGL4.10-1,31 kb	1.716.228	41.586	418.088,2	**89,7**	13,4
VII) pGL4.10-1,11 kb	1.799.574,3	57.166,5	315.072,7	**67,6**	4,5
VIII) pGL4.10-0,92 kb	1.837.957	50.902,3	360.421,3	**77,3**	7,8
IX) pGL4.10-0,63 kb	846.543	52.791	162.083,0	**34,8**	3,5
X) pGL4.10-0,41 kb	433.035,3	117.188,8	36.913,3	**7,9**	0,7

Eine zweite mögliche Erklärung für die schwankende *Renilla*-Expression ist eine unterschiedliche Transfektionseffizienz zu den verschiedenen Versuchszeitpunkten bzw. eine generell geringe und damit schnell schwankende Transfizierbarkeit der EP2102 Zellinie (Andrews et al., 1982). Die EP2102 Zellinie war in unserem Labor generell ungefähr 4 mal schlechter transfizierbar als die weibliche HEK (human embryonic kidney) 293T Zellinie (Graham et al., 1977), mit der Transfektionsraten zwischen 50 bis 60% erreicht wurden. Trotz optimierten Transfektionsprotokolls für die EP2102 Zellinie ist daher nicht auszuschließen, daß Unterschiede in der Transfektionseffizienz verschiedener Versuchsansätze, bei generell geringer Transfizierbarkeit der EP2102 Zellinie, die beobachteten Schwankungen erklären.
Die Konstruktion und die Sequenz der *DDX3X* Promoterkonstrukte könnten ebenfalls Einfluß auf die Ergebnisse dieser DLR-Assaysserie nehmen. Alle *DDX3X* Promoterkonstrukte beinhalten an der 3´Seite Sequenzen bis zur BAC-Position 30764, die bereits im kodierenden Bereich des *DDX3X* Exon1 liegt. Dieser Aufbau wurde gewählt, um möglichst den kompletten *DDX3X* Promoterbereich, inklusive des kartierten CpG-Island „CpG-P2" zu erfassen. D.h. aber auch, daß hierdurch vor dem Translationsstartcodon ATG, des *luc2*-Reportergens, drei weitere ATGs aus dem *DDX3X* Exon1 positioniert werden. Das native ATG für das DDX3X-Protein ist

dadurch 112 bp ou-of-frame stromaufwärts des *luc2* ATGs positioniert. Stromabwärts des nativen *DDX3X* ATGs sind weitere zwei ATGs in der *DDX3X* Exon1 Sequenz lokalisiert. Diese sind im pGL4.10 Vektor in-frame zum *luc2* ATG positioniert. Das erste ATG stromabwärts des nativen *DDX3X* ATGs ist 105 bp, das zweite ist 90 bp stromaufwärts des *luc2* ATGs positioniert. Bei einem Translationsstart an diesen *DDX3X* Exon1 ATGs, würde der *luc2*-Leserahmen N-terminal potentiell um 35 AS, bzw. 30 AS verlängert werden. Von den vier ATGs verfügt nur das *luc2* ATG über eine perfekte Kozak-Sequenz und somit über einen starken Sequenzkontext für die Proteintranslation. Eine Konkurrenz um den Translationsapparat in den EP2102 Zellen und mögliche Reinitiationsschritte am *luc2* ATG können aber nicht generell ausgeschlossen werden. Eine Überprüfung der potentiellen N-terminalen *luc2*-Proteinverlängerung, durch Translationsstart im den zwei *DDX3X* Exon1 ATGs, mit dem Protein-Domänen Analyse-Programm “Prosite” (Database of protein domains, families and functional sites), auf dem “ExPASy Proteomics Server”, ergab keine bekannten funktionellen Proteinmotive für die potentiellen 35 und 30 AS-Verlängerungen. Sollte diese Vektorkonstruktion Einfluß auf die Ergebnisse haben, sollten die vorgestellten Ergebnisse und das erhaltene *DDX3X* Promoter-Expressionsprofil im Vergleich der zehn Konstrukte dennoch aussagekräftig bleiben, da dieser Vektoraufbau für alle 10 *DDX3X* Promoterkonstrukte gilt.
Zusätzlich kommen stromaufwärts des nativen *DDX3X* ATGs, in den klonierten Promotersequenzen I bis X, weitere „upstream ATGs“ in unterschiedlicher Anzahl vor. Im längsten Promoterkonstrukt I (2,44 kb) kommen insgesamt 25 ATGs in der klonierten *DDX3X* Promotersequenz, stromaufwärts des *luc2* ATGs, vor. Im Anhang (Abb. A-5) ist die Sequenz des *DDX3X* Promoterkonstruktes I mit den potentiellen uATGs zu sehen.

Das in Abbildung 4-59 vorgestellte *DDX3X* Promoter-Expressionsprofil deutet darauf hin, daß die hier untersuchte *DDX3X* Promoterregion aus mehreren Aktivator- und/oder Repressordomänen aufgebaut ist. Die erste proximale *DDX3X* Aktivatordomäne ist mit den Konstrukten IX und X zu beschreiben. Diese Domäne kontrolliert wahrscheinlich die ubiquitäre *DDX3X* Expression. Die zweite, zentrale *DDX3X* Aktivatordomäne schließt die Keimzell-spezifische *DDX3X* T-TSS-III Position ein. Eine mögliche dritte, distale Aktivatordomäne ist im Bereich des W-Traktes Nr. 2 mit dem Konstrukt II darstellbar. Anhand der hier dargestellten Ergebnisse zur *DDX3X* Promoterregion, wird ein Promotermodell mit mindestens drei Aktivator-domänen vorgeschlagen, die in der EP2102 Zellinie funktionell sind. Danach würde die distale 5´Grenze der funktionellen *DDX3X* Promoterdomäne im Bereich zwischen Konstrukt I und II kartierbar, d.h. ~ 2,1 bis 2,4 kb stromaufwärts des *DDX3X* ATGs.

4.9.2. Kartierung partieller Funktionsbereiche im *DDX3X* Promoter

In der zweiten Promoter DLR-Testserie wurden die verschiedenen potentiell funktionellen *DDX3X* Promotersequenzabschnitte auf ihre Einzelaktivität im promoterlosen Vektor pGL4.10, sowie im Kontrollvektor pGL4.13 getrennt analysiert. Dabei galt es zu evaluieren, welche der *DDX3X* Promoterabschnitte eigene, inhärente Aktivatorqualität im pGL4.10 Vektor und welche davon nur potentielle Enhancer-, oder Repressorqualitäten im pGL4.13 Vektor, durch Modulation der endogenen SV40-Promoteraktivität, erkennen lassen.

Dazu wurden zwölf *DDX3X* Promoterfragmente mittels *HindIII* Restriktionsenzymschnittstellen („AAGCTT“) in die MCS der pGL4.10 und pGL4.13 Reportergenvektoren kloniert und anschließend auf Sequenzidentität und Orientierung überprüft. Generell wurden die Vektorkonstrukte selektioniert, in denen die klonierten *DDX3X* Promotersequenzen in nativer (+) Orientierung, wie auf dem Human X-Chromosom, vorkommen. In Abb. 4-60 sind die partiellen *DDX3X* Promotersequenzen schematisch in Bezug zu den bisher kartierten potentiellen Promoterelementen und in Bezug zu den Promoterkonstrukten der ersten DLR-Serie dargestellt. Für die zweite DLR-Serie wurden präferentiell Sequenzen, wie z.B. R- und Y-Trakt, gewählt, die in Kapitel 4.6. eine X-spezifische Evolution zeigten. In Tab. 4-29 sind alle partiellen *DDX3X* Promoterkonstrukte mit Position im BAC RP13-13A3 und Insertlänge aufgelistet.

Tab. 4-29: Liste der partiellen *DDX3X* Promoterkonstrukte.
Die partiellen *DDX3X* Promotersequenzen sind entsprechend Abb. 4-60 mit Kleinbuchstaben von proximal nach distal in der *DDX3X* Promoterregion durchnummeriert.

Nummerierung	Konstruktbezeichnung	Position im BAC RP13-13A3	Länge des klonierten Fragmentes
a	AG-Box	30138 - 30378	241 bp
b	AG-Box ex(tended)	30138 - 30564	427 bp
c	POZ1	29198 – 29888	691 bp
d	POZ2	29460 - 29888	429 bp
e	POZ3	29658 - 29888	231 bp
f	TC-Box	29785 - 29981	197 bp
g	CpG1	28443 - 29863	1421 bp
h	CpG2	28779 – 29863	1085 bp
i	CpG3	28915 - 29863	949 bp
j	W-Box	28443 – 28800	358 bp
k	W-Box ex(tended)	28443 – 28943	501 bp
l	Aktivator/Repressor	28779 - 28970	192 bp
IIΔ	2,13 kb Δ POZ	28638 - 29104 + 30034 - 30764	1198 bp (929 bp Δ)

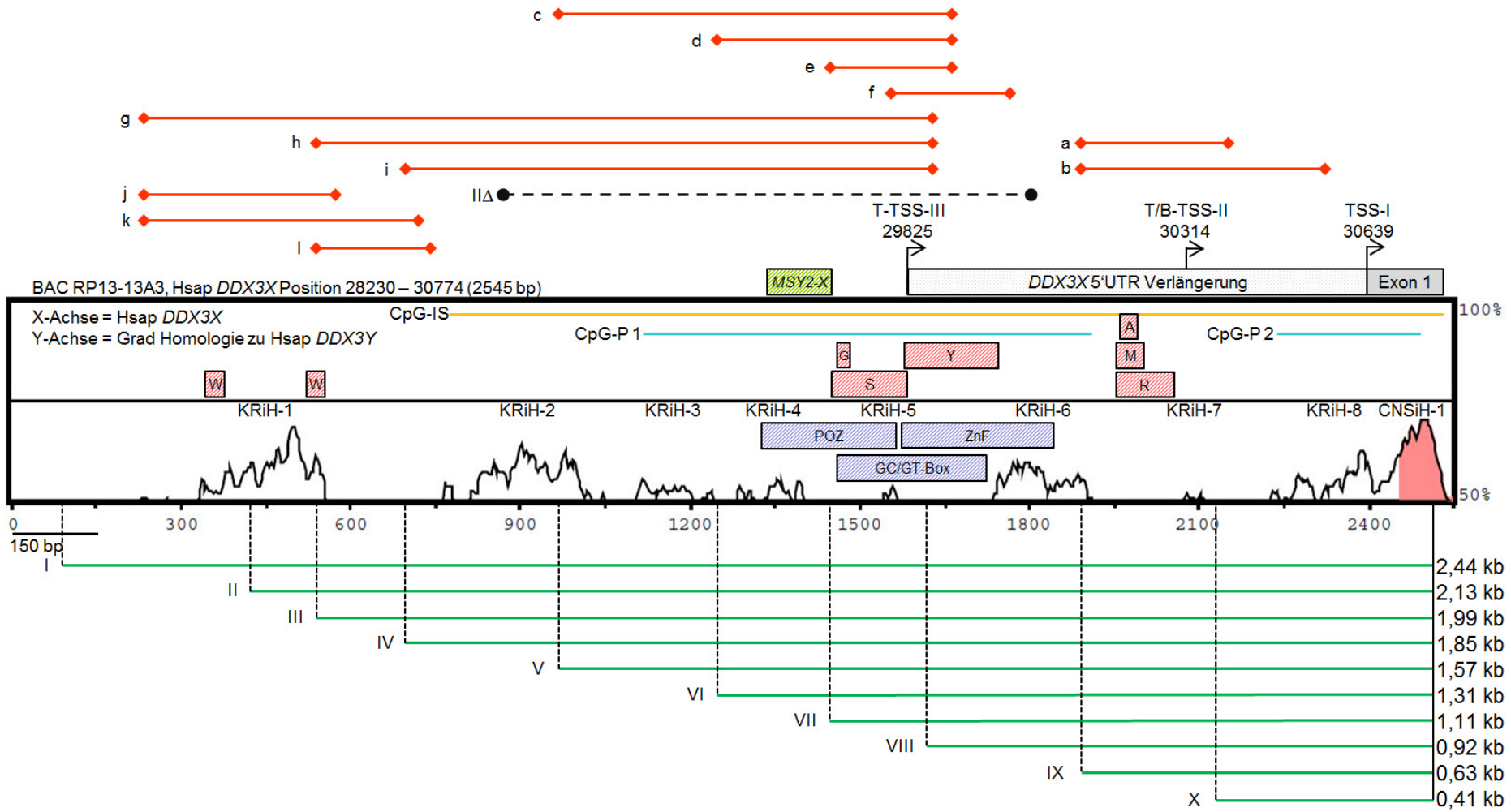


Abb. 4-60: Graphische Darstellung der partiellen *DDX3X* Promoterkonstrukte.

Die Graphik basiert auf den Ergebnissen der *DDX3X* DLR-Analyse in Kapitel 4.9.1 und der *in-silico* Promoteranalyse in Kapitel 4.6. Die zwölf partiellen *DDX3X* Promoterkonstrukte sind durch Kleinbuchstaben gekennzeichnet, die in Tabelle 4-29 wiederzufinden sind. Konstrukt IIΔ bezieht sich auf das *DDX3X* Promoterkonstrukt II. Positionsangaben beziehen sich auf den Human BAC-Klon RP13-13A3.

Als erster *DDX3X* Promoterabschnitt wird mit den Inserts a (AG-Box) und b (AG-Box extended) der Bereich um die *DDX3X* T/B-TSS-II Position betrachtet (Abb. 4-60). Beide Fragmente beinhalten die AG-reichen Sequenzabschnitte direkt stromaufwärts der T/B-TSS-II Position und die Initiationsstelle selbst. Für das Insert a, welches in der X-spezifischen Sequenz zwischen „KRiH-6“ und „KRiH-7“ liegt, konnte im pGL4.13 Vektor eine 5,3fache Steigerung der rel. Luziferaseaktivität gemessen werden (Abb. 4-61, Tab. 4-30). Der gemessene Firefly-Luziferasewert war mit 29,3 Mio. „relative light units“ (RLUs) der höchste Wert der gesamten DLR-Testreihe. Trotz des sehr starken Effektes im pGL4.13 Vektor, konnte für das Insert a im promoterlosen Vektor pGL4.10 nur eine sehr moderate Luziferaseaktivität von ~10%, und somit nur geringe Eigenaktivität, festgestellt werden.

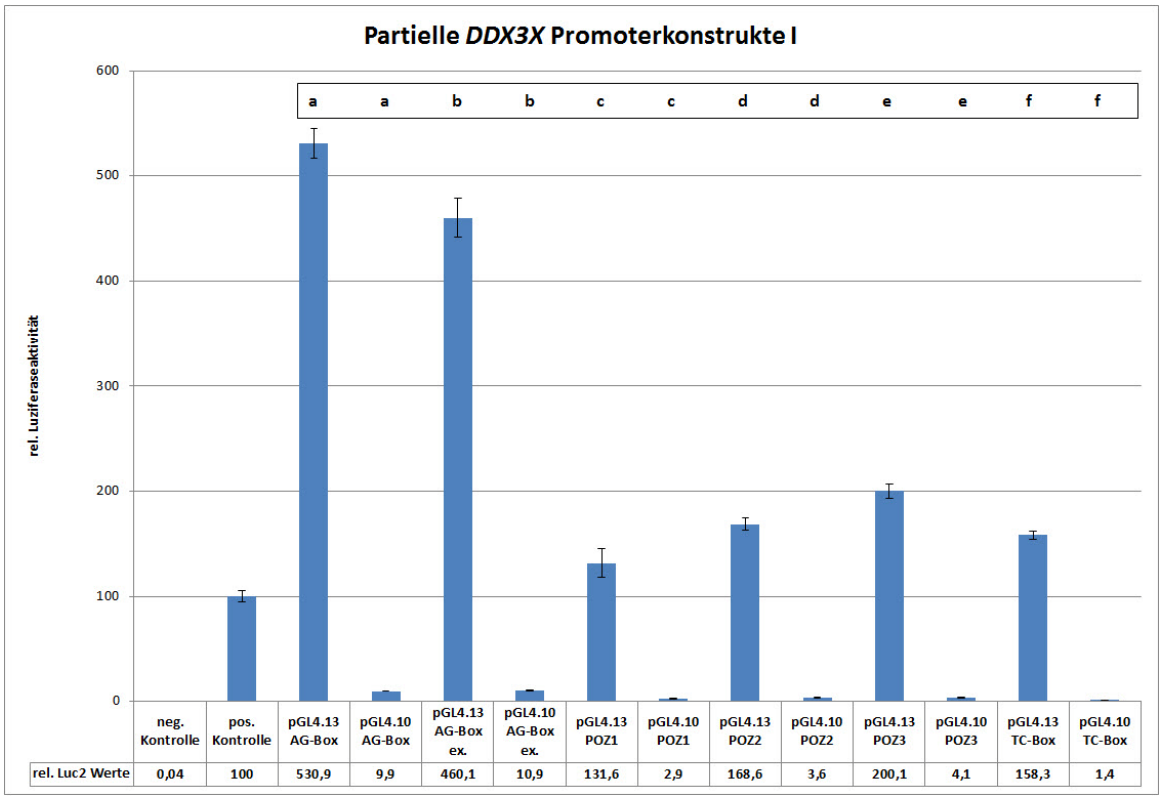


Abb. 4-61: Auswertung der partiellen *DDX3X* Promoterkonstrukte I.
Experimentelle Vektoren sind der endogen promoterlose Vektor pGL4.10 und der pGL4.13 Vektor. Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3X* Promoter-Sequenzabschnitte a bis f dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-30 aufgeführt.

Insert b (AG-Box extended) ist, im Vergleich zu Insert a, 3´ um 186 Nukleotide verlängert. Mit Insert b konnte ebenfalls eine sehr starke absolute und relative Luziferaseaktivität im Vergleich zur Positivkontrolle gemessen werden. Der 4,6fache Aktivitätsanstieg ist nur im pGL4.13 und nicht im pGL4.10 Vektor festzustellen. Die rel. Luziferaseaktivität im pGL4.10 Vektor beträgt ~ 11%. Beide Inserts a und b zeigen also einen starken Enhancereffekt auf den endogenen

SV40-Promoter im pGL4.13 Vektor, verfügen aber über nur geringe Eigenaktivität als potentielle Promotersequenzen im pGL4.10 Vektor.

Der Vergleich mit dem pGL4.10 *Luc*-Expressionsverhalten des Promoterkonstrukt IX der ersten DLR-Serie (Abb. 4-59) zeigt, daß eine weitere Verlängerung 3´ um 200 Nukleotide bis ins *DDX3X* Exon1, für eine Verstärkung der Promoteraktivität auf ~ 35% genügt. Zusammen mit dem Ergebnis von Konstrukt X, ist dies somit der erste Hinweis, daß der *DDX3X* Promoter-kontext, sowie zusammenhängende Promoterbereiche für eine eigene Promoteraktivität dieser proximalen *DDX3X* Promoterdomäne im Bereich der TSS-I Position entscheidend sind.

Tab. 4-30: Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion I DLR-Analyse.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
a) pGL4.13 AG-Box	29.318.538,7	118.555,3	2.474.474,1	**530,9**	14
a) pGL4.10 AG-Box	963.451	208.376	46.220,8	**9,9**	0,2
b) pGL4.13 AG-Box ex.	26.033.354,7	121.729	2.144.705,4	**460,1**	18,4
b) pGL4.10 AG-Box ex.	721.528	142.393,7	50.823,6	**10,9**	0,5
c) pGL4.13 POZ1	9.259.430,6	152.470,6	613.440,0	**131,6**	13,5
c) pGL4.10 POZ1	178.966,8	137.038,8	13.379,6	**2,9**	0,5
d) pGL4.13 POZ2	11.588.522	147.534,3	785.903,1	**168,6**	5,8
d) pGL4.10 POZ2	339.083	204.689	16.581,7	**3,6**	0,4
e) pGL4.13 POZ3	17.177.352,7	184.442,3	932.563,7	**200,1**	6,8
e) pGL4.10 POZ3	502190	264.752,8	19.059,5	**4,1**	0,5
f) pGL4.13 TC-Box	26.234.638,7	355.324,3	738.000,0	**158,3**	3,7
f) pGL4.10 TC-Box	507.656,4	787.554,4	6.424,8	**1,4**	0,06

Mit den Inserts c bis f wurde die potentielle Promoterregion um und stromaufwärts der *DDX3X* T-TSS-III Position untersucht. Für das Konstrukt c (POZ1) konnte im pGL4.13 Vektor eine 1,32fache Verstärkung der Luziferaseaktivität gemessen werden. Derselbe Promoterabschnitt im promoterlosen Vektor pGL4.10 ergab nur eine minimale Aktivität von 2,9% (Abb. 4-61). D.h. trotz Anwesenheit der *DDX3X* T-TSS-III Position und der flankierenden Promoterelemente in diesem Konstrukt c, ist dieser klonierte Promoter-Subbereich alleine nicht ausreichend für eine selbstständige Aktivierung des Reportergens.

Die schrittweise Eingrenzung dieser *DDX3X* Promoterregion mit den Konstrukten d (POZ2) und e (POZ3) hat eine 1,7 und 2fache Aktivitätssteigerung für Konstrukt d bzw. für Konstrukt e, in Bezug zum pGL4.13 Vektor, zur Folge (Abb. 4-61). Im Vergleich zu Konstrukt c entspricht dies einer 1,3fachen Steigerung für Konstrukt d, bzw. einem 1,5fachen Anstieg für e.
Dieselben Sequenzabschnitte im promoterlosen pGL4.10 Vektor zeigen nur sehr geringe rel. Luziferaseaktivitäten von 3,6% (d) bzw. 4,1% (e). Die Verkleinerung der klonierten *DDX3X* Promoterabschnitte in den Konstrukten von c bis e führte also zu einer Steigerung der positiv regulativen Wirkung („Enhancer-Effekt“), bis zu Faktor 2 für das Insert e, im pGL4.13 Vektor. Für diese zentrale *DDX3X* Promoterdomäne wird, im Vergleich zu den pGL4.10 Konstrukten V, VI und VII der ersten DLR-Serie, die in denselben BAC-Positionen wie die Inserts c, d und e beginnen, deutlich, daß der proximale *DDX3X* Promoterabschnitt, der in den Konstrukten c, d und e nicht enthalten ist, benötigt wird, um einen *DDX3X* Promoterkontext mit starkem endogenem Aktivierungspotential zu etablieren.

Das Konstrukt f (TC-Box) beinhaltet die Pyrimidin-reiche Region, in die die *DDX3X* T-TSS-III Position eingebettet ist. Mit diesem Insert f soll ein möglicher regulativer Effekt untersucht werden, der für TC-Boxen beschrieben wurde (siehe Kapitel 4.6.2).

In Abb. 4-61 ist zu sehen, daß auch dieser Promoterabschnitt f über keine eigene Promoteraktivität im pGL4.10 Vektor verfügt. Im pGL4.13 Vektor ist eine 1,58fache Verstärkung der endogenen SV40-Promoteraktivität mit Insert f zu messen. Aus dem Vergleich der Inserts e und f, die in knapp 100 bp überlappen, läßt sich für die EP2102 Zellinie schlußfolgern, daß die längere Pyrimidin-reiche Sequenz in Insert f nicht zu einer weiteren Verstärkung der potentiellen Enhancerfunktion, in diesem Versuchsaufbau, führt.

Die Untersuchungen zur Funktion des distalen *DDX3X* Promoterbereiches erfolgen mit den Konstrukten g bis l (Abb. 4-60; Abb. 4-62). Das Insert g (CpG1) umfaßt alle potentiellen distalen Promoterelemente bis einschließlich der *DDX3X* T-TSS-III Position (Abb. 4-60). Im pGL4.10 Vektor konnte mit Insert g nur eine geringe rel. Luziferaseaktivität von 4,3% festgestellt werden (Abb. 4-62, Tab. 4-31). Dieser Sequenzabschnitt weist somit ebenfalls keine endogene Promoteraktivität auf. Im pGL4.13 Vektor führte dieses Insert e zu einer Reduktion der rel. Luziferaseaktivität auf 54,9% und ist damit der erste partielle *DDX3X* Promoterabschnitt, der in nativer Orientierung zu einem Aktivitätsverlust des SV40-Promoters führte. Für das am 5´Ende 336 Nukleotide kürzere Insert h (CpG2) konnte in pGL4.13 eine 2,4fache Steigerung der rel. Luziferaseaktivität gemessen werden. Im promoterlosen pGL4.10 konnte eine rel. Luziferaseaktivität von 16,7% festgestellt werden. Das Insert h verfügt über eine eigene moderate Promoteraktivität.

Das Insert i (CpG3), welches im Vergleich zu Insert h nochmals um 136 Nukleotide verkürzt wurde, zeigt im pGL4.13 Vektor keine Aktivitätsveränderung gegenüber dem pGL4.13 Original-Vektor ohne Insert (Abb. 4-62). Im promoterlosen Vektor pGL4.10 ist mit Insert i nur eine sehr geringe rel. Luziferaseaktivität von 4% zu messen.

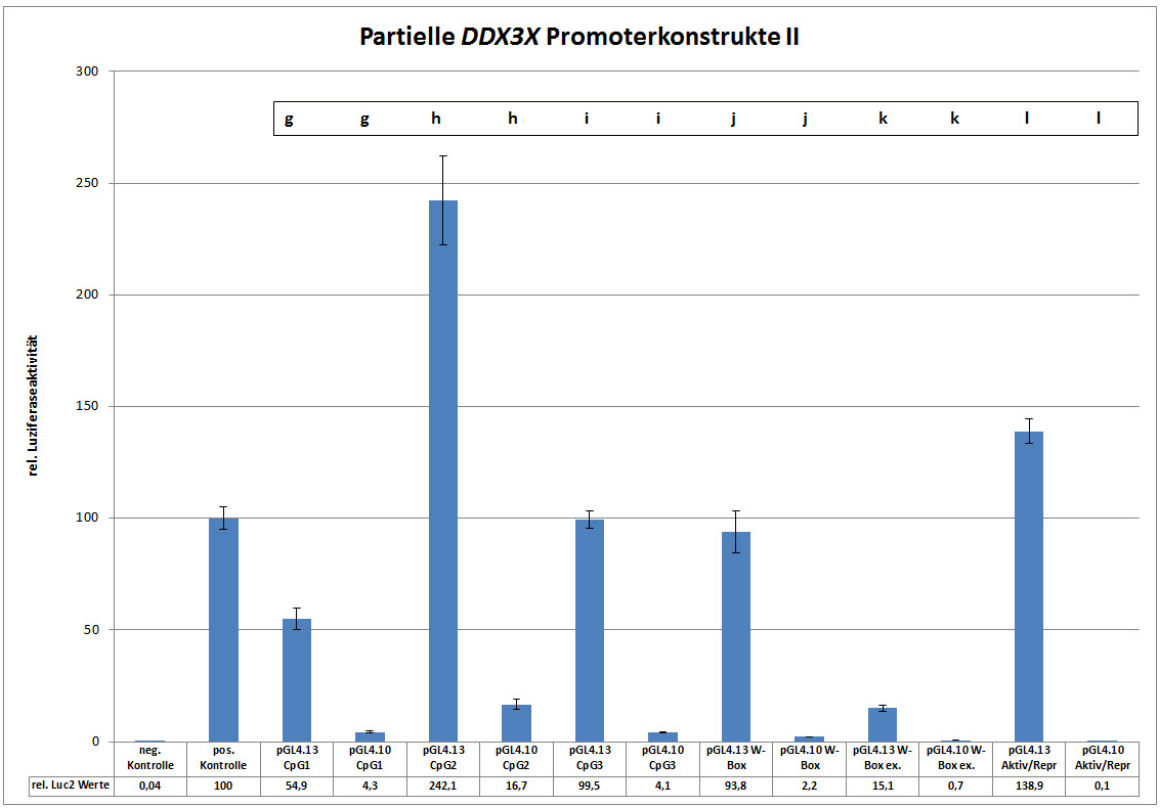


Abb. 4-62: Auswertung der partiellen *DDX3X* Promoterkonstrukte II.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der distalen *DDX3X* Promoter-Sequenzabschnitte g bis l dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-31 aufgeführt.

In dem 136 bp längeren Sequenzabschnitt in Insert h sollten demnach, im Vergleich zu Insert i, positiv regulierende Elemente vorkommen. Im 5´ Bereich von Insert g sollten dagegen reprimierende Elemente liegen, die in diesem Versuchsaufbau zu einer Repression der SV40 Promoteraktivität führten.

Der betreffende 5´ Sequenzabschnitt aus Insert g wird daher mit dem Insert j (W-Box auf Grund der zwei kartierten W-Trakte genannt) isoliert betrachtet (Abb. 4-60). In Abb. 4-62 und Tab. 4-31 ist zu sehen, daß Insert j im pGL4.13 Vektor kaum einen Effekt auf die rel. Luziferaseaktivität des Vektors ausübt. Die rel. Luziferaseaktivität von ~ 94% zeigt keinen signifikanten Unterschied zur Positivkontrolle. Der reprimierende Effekt der im Insert g (CpG1), im Vergleich zu Insert h, festgestellt wurde, konnte mit diesem partiellen Insert j (W-Box) nicht dargestellt werden. Im promoterlosen Vektor zeigt auch dieser Sequenzabschnitt keine eigene Luziferase-aktivität.

Die Verlängerung dieser Sequenz j am 3´Ende um den potentiell aktivierenden Bereich aus Insert h führte zur Etablierung von Insert k (W-Box extended). Im pGL4.13 Vektor zeigte sich ein deutlicher Repressoreffekt. Es konnte eine 6,7fache Aktivitätsreduktion auf 15,1% rel. Luziferaseaktivität gemessen werden. Eine endogene Promoteraktivität im pGL4.10 Vektor konnte mit Insert k nicht festgestellt werden.

Tab. 4-31: Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion II DLR-Analyse.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
g) pGL4.13 CpG1	6.394.019,2	251.404,6	255.912,0	**54,9**	4,8
g) pGL4.10 CpG1	515.470	261.116	19.954,1	**4,3**	0,5
h) pGL4.13 CpG2	16.219.111,3	144.202,3	1.128.524	**242,1**	19,9
h) pGL4.10 CpG2	1.297.912,5	52.855	77.695,7	**16,7**	2,2
i) pGL4.13 CpG3	9.271.537	200.329,8	463.582,3	**99,5**	3,9
i) pGL4.10 CpG3	594.529,5	314.338	18.932	**4,1**	0,1
j) pGL4.13 W-Box	1.769.168,8	52.335,3	437.016,1	**93,8**	9,4
j) pGL4.10 W-Box	87.305,5	84.929,3	10.282,0	**2,2**	0,05
k) pGL4.13 W-Box ex.	497.378,7	71.160,7	70.347,3	**15,1**	1,4
k) pGL4.10 W-Box ex.	21.861	63.777,5	3.470,8	**0,7**	0,09
l) pGL4.13 Aktivator/Repressor	11.691.846	180.675,3	647.541,0	**138,9**	5,5
l) pGL4.10 Aktivator/Repressor	4.923,4	198.423,6	251,8	**0,1**	0,01

Für die Promoter-Subkonstrukte der distalen *DDX3X* Promoterdomäne, im Bereich der KRiH-1, scheinen in diesem Versuchsaufbau eindeutig Positionseffekte zu existieren. Der ~ 160 bp lange *DDX3X* Promoterabschnitt, der am 5´Ende in Insert h zu einem aktivierenden Effekt führt, vermittelt am 3´Ende in Insert k einen reprimierenden Effekt. Unter Umständen führt die Länge der klonierten Promoterfragmente, durch Ausbildung räumlich unterschiedlicher Promoter-Sekundärstrukturen, zu der beobachteten Modulation der Luziferaseaktivitäten.

Um den offensichtlich regulativ ambivalenten Effekt dieser Sequenz (zwischen h 5´Ende und k 3´Ende) gezielt untersuchen zu können, wurde das 192 bp lange Insert l (Aktivator/Repressor) generiert. Im pGL.4.13 Vektor konnte für Insert l eine verstärkende Wirkung auf den bestehenden SV40-Promoter festgestellt werden (Abb. 4-62). Die rel. Luziferaseaktivität stieg auf ~ 139%. Eine eigenständige Promoteraktivität im promoterlosen Vektor pGL4.10 war nicht

messbar. Die Funktion dieses *DDX3X* Promoter-Subabschnittes ist also abhängig vom Kontext der umliegenden Sequenzen.

Da für die klonierten *DDX3X* Promoterabschnitte im Bereich von Insert I starke Positionseffekte erkennbar waren, sollte auch überprüft werden, ob zusätzlich Orientierungseffekte auf das *Luc*-Expressionspotential vorlagen. Dazu wurde das Insert I in reverser Orientierung (- Ori) in die pGL4.13 und pGL4.10 Vektoren eingeführt.

In reverser Orientierung führte Insert I zu einer 1,92fachen Aktivitätssteigerung der rel. Luziferaseaktivität (Abb. 4-63, Tab. 4-32). Im Vergleich zur nativen Orientierung eine Steigerung um Faktor 1,38. Im promoterlosen Vektor war nach Orientierungsumkehr kein Effekt meßbar. D.h. auch nach Umkehr der Orientierung bleibt die endogene Promoteraktivität dieser Region bei null in diesem Versuchsaufbau. Insert I zeigt deutlich zwei Effekte in dieser DLR-Assay Serie. Erstens konnte ein Positionseffekt und zweitens ein Orientierungseffekt beobachtet werden. Die unterschiedlich starken Luziferaseaktivitäten der Promoterkonstrukte II bis V könnten, wie in Kapitel 4.9.1 vorgeschlagen auf einer dritten distalen Aktivatordomäne beruhen, oder auf weiteren Positions- und Orientierungseffekten bestimmter Promoterelemente.

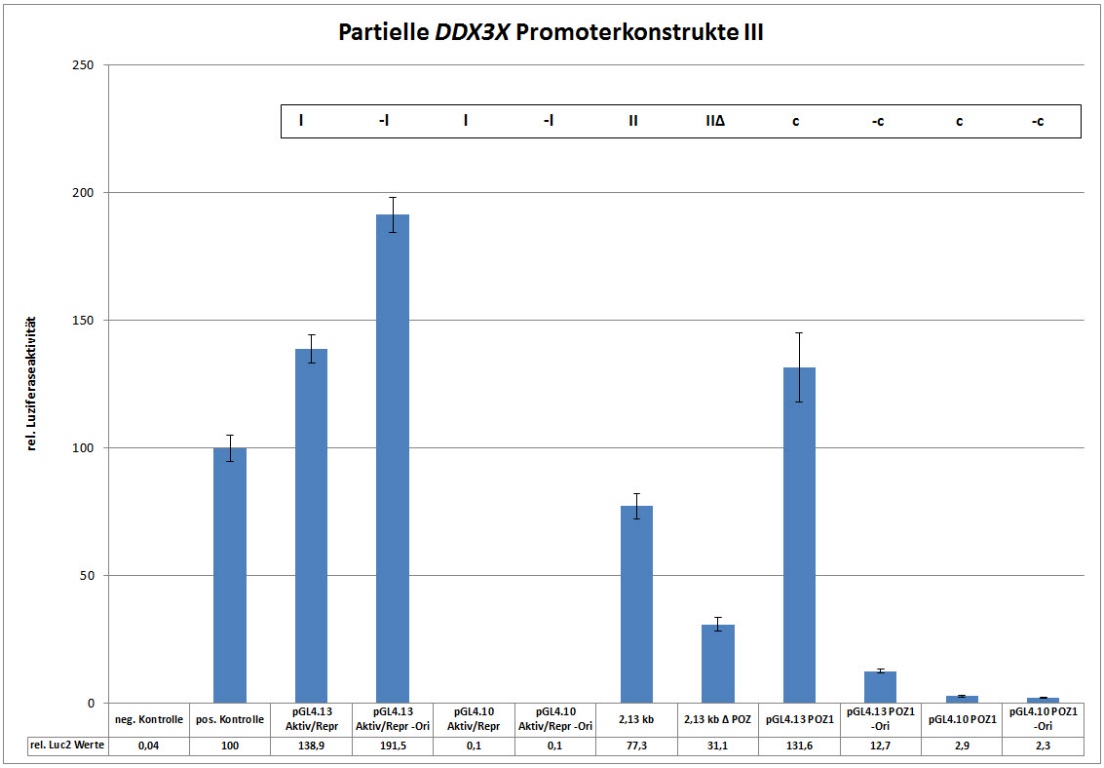


Abb. 4-63: Auswertung der partiellen *DDX3X* Promoterkonstrukte III.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3X* Promoter-Sequenzabschnitte I und c in reverser Orientierung, sowie dem Deletionskonstrukt IIΔ dargestellt. Die relativen und gemittelten Luziferase -werte sowie die prozentualen Standardabweichungen sind in Tab. 4-32 aufgeführt.

In Kapitel 4.6.3.2. konnte eine auffällige Clusterung von potentiellen ZF5-TFBS zwischen BAC-Position 29.552 und 29.812 (BAC-Klon RP13-13A3) im *DDX3X* Promoter gezeigt werden (Abb. 4-60). TFs der POZ-Familie werden häufig als transkriptionelle Repressoren beschrieben (Collins et al., 2001; Kelly & Daniel, 2006, Costoya, 2007). Um den möglichen Einfluß dieser Region im Kontext einer längeren *DDX3X* Promotersequenz untersuchen zu können, wurde ein Deletionskonstrukt, durch Verwendung zwei interner Restriktionsenzymschnittstellen (*EcoRI*; *PmlI*) in der *DDX3X* Promotersequenz, generiert. Als Template wurde das *DDX3X* Promoterkonstrukt II (2,13 kb) gewählt, da dieses wahrscheinlich alle drei vorgeschlagenen *DDX3X* Promoterdomänen vollständig enthält. Im Deletionskonstrukt IIΔ wurden zwischen den endogenen *EcoRI*- (BAC-Pos. 29105) und *PmlI*-Schnittstellen (BAC-Pos. 30033) 929 bp durch Restriktionsenzymverdau deletiert und anschließend durch „blunt-end Ligation“ wieder geschlossen. Die Deletion betrifft die Promoterbereiche, die mit den partiellen Inserts c bis f untersucht wurden. Für das Deletionskonstrukt IIΔ (pGL4.10-2,13 kbΔ POZ) konnte eine 2,5fache Reduktion der rel. Luziferaseaktivität im Vergleich zum Original-Konstrukt II festgestellt werden (Abb. 4-63). Entgegen der ursprünglichen Annahme, daß diese POZ-TF Clusterung im nativen Promoterkontext zu einer transkriptionellen Repression führen kann, zeigte sich in diesem Experiment, erst nach POZ-Deletion, eine Repression dieser *DDX3X* Promoterregion. Die Promoter-Subkonstrukte c, d und f zeigen dies ebenfalls.

Tab. 4-32: Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion III DLR-Analyse.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
I) pGL4.13 Aktiv/Repr -Ori	15.553.679,3	174.686	892.438,3	**191,5**	6,9
I) pGL4.10 Aktiv/Repr -Ori	6.026,5	254.607	237,9	**0,1**	0,003
IIΔ) pGL4.10-2,13 kb Δ POZ	1.319.038,2	89.772,4	144.767,3	**31,1**	2,5
c) pGL4.13 POZ1 -Ori	799.841,3	134.786,5	59.348,7	**12,7**	0,7
c) pGL4.10 POZ1 -Ori	65.880,3	61.352,7	10.905,4	**2,3**	0,3

Um zu evaluieren, ob für diesen *DDX3X* Promoterabschnitt die Orientierung von Bedeutung ist, wurde das Insert c in reverser Orientierung in die pGL4.13 und pGL4.10 Vektoren kloniert. Im pGL4.13 Vektor konnte eine 10,4fache Reduktion auf 12,7% rel. Luziferaseaktivität, im Vergleich zu Insert c in nativer Orientierung, festgestellt werden. Im promoterlosen Vektor blieb die rel. Luziferaseaktivität unverändert sehr gering, so daß auch für Insert c in -Ori keine endogene Promoteraktivität feststellbar ist. In diesem Versuchsaufbau war es also wichtig, daß

die *DDX3X* Promoterelemente im Bereich der T-TSS-III Position (Abb. 4-60) in nativer Orientierung im Reportergenvektor analysiert wurden.

Diese zweite DLR-Assay Serie zeigt deutlich, daß Position und Orientierung, also das Zusammenspiel aller *DDX3X* Promoterelemente im richtigen Kontext, von Bedeutung für die Funktionalität der gesamten *DDX3X* Promoterdomäne ist. Auch kann eine Spezifität der getesteten regulativen Elemente für den nativen *DDX3X* Promoter vorliegen, wodurch ganz allgemein die inhärente Funktionalität in einem Versuchsaufbau mit „Fremdpromoter“ nicht korrekt erfaßt werden kann.

4.10. Funktionelle Analysen zur repetitiven *MSY2* Minisatellitenrepeat-sequenz in der *DDX3Y* Promoterdomäne und zur *MSY2-X* Sequenz in der *DDX3X* Promoterdomäne

Die Entstehung einer neuen *DDX3Y* Core-Promoterregion in der *MSY2* Minisatellitensequenz der catarrhinen Primaten, im Vergleich zur homologen Sequenz *MSY2-X* vor dem *DDX3X* Exon1 (siehe auch Kapitel 4.3., Kapitel 4.6.), ließ vermuten, daß der Y-chromosomale *MSY2*-Repeat ein eigenständiges Promoterelement mit Aktivatorqualität darstellt.

Minisatelliten gelten ganz allgemein als regulative Elemente, die bereits für andere Gene beschrieben wurden (Vergnaud & Denoeud, 2000; Michelhaugh et al., 2001; Williams et al., 2001; Inoue-Murayama et al., 2008; Lee et al., 2008). Minisatellitenrepeats werden auch VNTRs („Variable Number of Tandem Repeats“) auf Grund ihres polymorphen Charakters in Bezug auf die Anzahl der Repeateinheiten genannt (Buard & Jeffreys, 1997; Jeffreys et al., 1997; Bérard & Rivals, 2003). Für die Entstehung solcher Tandem-Repeats, in denen die duplizierte Repeateinheit neben dem Original-Sequenzblock ins Genom eingefügt wird, werden ein Verrutschen der DNA-Polymerase während der Replikation („Slippage“) und ungleiches Crossing-over vermutet (Bérard & Rivals, 2003). Vergnaud & Denoeud (2000) beschreiben die Basenpaarzusammensetzung von Minisatelliten generell als (G,C)-reich, es existieren aber ebenfalls sehr AT-reiche Minisatelliten, für die S/MAR-ähnliche Funktionen beschrieben werden konnten (Jackson et al., 2003). Die Minisatelliten-Instabilitäten, die chromosomale Brüche hervorrufen können, scheinen nicht spezifisch für (A,T)- oder (C,G)-reiche Repeats zu sein. Solche Bruchereignisse sind überwiegend in der Keimbahn während der meiotischen Rekombinationen zu erwarten (Jeffreys et al., 1998; Bois & Jeffreys, 1999; Buard et al., 2002). Da die Basenzusammensetzung eines Minisatelliten als ein Charakteristikum gilt, wurde der (G,C)-Gehalt der *MSY2* Repeateinheiten mit dem der *MSY2-X* Sequenz verglichen. Der (G,C)-

Gehalt beträgt für *MSY2-1* 46,5%, 43,6% für *MSY2-2* und für *MSY2-3/4* beträgt der Wert 45,4%. Die *MSY2-X* Sequenz hat einen (G,C)-Gehalt von 65,4%. D.h., die Sequenzen divergieren im (G,C)-Gehalt um +/- 20 Prozentpunkte. Der deutliche Unterschied in der Basenzusammensetzung könnte also Einfluß auf eine potentielle, regulative Funktion dieser Sequenzen nehmen, da für (A,T)- und (G,C)-reiche Minisatelliten verschiedene Funktionen beschrieben wurden.

In Abb. 4-64 ist ein vergleichendes ClustalW-Sequenzalignment zwischen den *MSY2* Repeateinheiten und *MSY2-X* zu sehen. Die Homologie zwischen den vier *MSY2* Repeateinheiten und der 107 bp langen X-chromosomalen *MSY2-X* Sequenz konnte durch die Konserviertheit der potentiellen SOX bzw. DMRT Transkriptionsfaktor-Bindestelle erkannt werden (Kapitel 4.6.3.1.). Per Sequenzvergleich konnte auch der *MSY2* Repeat 1 (BAC 475I1, Position 53.854 - 53.954) als die ursprünglich homologe Repeateinheit zu *MSY2-X* (BAC RP13-13A3, Pos. 29.570 - 29.676) definiert werden (Abb. 4-64). Die Sequenzhomologie zwischen *MSY2-1* und *MSY2-X* beträgt 63,4%. Zwischen den evolutionär später auf dem Y-Chromosom duplizierten Repeateinheiten 2, 3 und 4, welche identisch mit *MSY2* Repeat 3 ist, betragen die Sequenzhomologien 59,8% bzw. 57,6% (Abb. 4-64). Für *MSY2-X* konnte in keiner der untersuchten Spezies eine Duplikation des Sequenzintervalls stromaufwärts zum *DDX3X* Gen identifiziert werden.

Der *MSY2* Minisatellitenrepeat (*DYS440*) wurde erstmals von Bao et al. (2000) für Mensch (4 Repeateinheiten), Schimpanse (2 Repeateinheiten) und Orang-Utan (2 Repeateinheiten) beschrieben. In Kapitel 4.4.1 konnte zudem gezeigt werden, daß für *Macaca* und *Callithrix* je zwei Y-chromosomale *MSY2* Repeateinheiten existieren (siehe auch Rauschendorf et al., 2011). Die *MSY2* Sequenzhomologien zwischen den nicht-Human Primaten und dem Human *MSY2* Repeat betragen zwischen 78% und 100%. Die identifizierten *MSY2* Sequenzen in Rind (109 bp) und Maus (97 bp) weisen Sequenzhomologien von 70,3% bzw. 70,1% zum Human *MSY2* Repeat 1 auf (Alignment siehe Kapitel 4.4.2).
Die Homologiewerte zwischen der Human *MSY2-X* Sequenz und den homologen Sequenzen in nicht-Human Primaten betragen in Schimpanse 98,1% (107 bp), in *Macaca* 94,4% (107 bp) und in *Callithrix* ebenfalls 94,4% (107 bp). Die entsprechenden *MSY2-X* Sequenzen in Rind und Maus weisen Homologiewerte von 69,8% (106 bp) bzw. von 68,2% (114 bp) auf.
Die ermittelten *MSY2* und *MSY2-X* Sequenzdivergenzen, weisen eine abnehmende Homologie von Mensch bis Maus, entsprechend, den evolutionären Abständen der hier untersuchten sechs Säugerspezies auf. Eine Ausnahme ist nur für das Rind feststellbar, welches stets eine höhere Homologie, trotz größeren evolutionären Abstandes zum Menschen, als die Maus aufweist (siehe auch Kapitel 4.4.2., 4.5.3.).

Abb. 4-64: Vergleich der Human *DDX3Y MSY2* und *DDX3X MSY2-X* Sequenzen.

DDX3Y MSY2 Minisatelliten-Sequenzen aus BAC 475I1. *MSY2-1* Position 53.854 - 53.954 (101 bp), *MSY2-2* Position 53.955 - 54.064 (110 bp), *MSY2-3* Position 54.065 - 54.163 (99 bp), *MSY2-4* Position 54.164 - 54.262 (99 bp). *DDX3X MSY2-X* Sequenz aus BAC RP13-13A3, Position 29.570 - 29.676 (107 bp). Der Sequenzvergleich mit Repeateinheit *MSY2-4* ist nicht gezeigt, da dieser identisch ist mit Repeat Nr. 3. Die Sequenzidentität zwischen *MSY2-1* und *MSY2-X* beträgt 63,4%, zwischen *MSY2-2* und *MSY2-X* 59,8% und zwischen *MSY2-3/4* und *MSY2-X* jeweils 57,6%. Der Homologiewerte wurden nicht mit „Gap-Penalties“, sondern mit der Anzahl an Nukleotididentitäten berechnet.

Das hexamere Palindrom „CCTAGG“ in der *MSY2* Sequenz ist blau markiert. Die Verdopplung in *MSY2-2/3/4* ist hellblau markiert. Das entsprechende Hexamer „CCGGGG“ ist in der *MSY2-X* Sequenz lilafarben hervorgehoben. In *MSY2-1* liegt eine potentielle TFBS für die Faktoren TR2 und TR4 („Testicular Receptor“ 2 & 4) aus der Familie der Steroidhormon-Rezeptoren („Nuclear Receptor Subfamily 2, Group C“) in dem Palindrom. In den *MSY2* Repeateinheiten 2, 3 und 4 wird diese TFBS durch die Verdopplung des Palindromes eliminiert. Die Hexamersequenz „CCGGGG“ in *MSY2-X* ist ebenfalls eine potentielle TFBS für einen nukleären Hormonrezeptor („Nuclear Receptor Subfamily 1, Group I). Der vorhergesagte TF ist VDR als Heterodimer mit RXR (Heterodimer aus „Vitamin D3 Rezeptor“ & „Retinoid X Rezeptor“). Eine konservierte TFBS zwischen Human *DDX3Y* und *DDX3X* wird von Genomatix „DiAlign TF“ nicht für das Palindrom angegeben.

Das verdoppelte Oktamer „ATTGTTTC“ ist in den *MSY2* Sequenzen grün bzw. hellgrün markiert und diente zur Identifizierung der homologen *MSY2-X* Sequenz. Das konservierte Oktamer in der *MSY2-X* Sequenz ist ebenfalls hellgrün hervorgehoben. Das Derivat des ersten Oktameres in *MSY2-X* ist orangefarben markiert. Als potentielle TFs für diese Sequenz wurden bereits in Kapitel 4.6.3.1. Faktoren der SOX und DMRT TF-Familie vorgestellt. In den *DDX3Y MSY2* Repeats sind jeweils zwei SOX/DMRT TFBS vorhanden.

4.10.1. Funktionelle Analyse des *DDX3Y MSY2* Minisatellitenrepeats

Die Funktionsanalyse des *DDX3Y MSY2* Repeats ist eine Fortführung und Erweiterung des Projektes zur Charakterisierung der *DDX3Y* Promoterdomäne, welche in unserem Labor bisher durch Hanstein (2004) und Fröhlich (2008) durchgeführt wurde.

Da bekannt ist, daß die Klonierung repetitiver DNA zu Instabilität und Rearrangements der Sequenzen führen kann (Neil et al., 1990; Glover & Hames, 1995), wurde entweder der Forwardprimer oder der Reverseprimer außerhalb der Repeatsequenz positioniert, um die Klonierbarkeit der Repeats in die pGL4.10 und pGL4.13 Vektoren sicherzustellen. Insgesamt wurden neun unterschiedliche Inserts generiert, die sich in der Anzahl der *MSY2* Repeat-einheiten unterscheiden (Abb. 4-65, Tab. 4-33). Es wurden drei Inserts mit fixer 5´Position in BAC-Position 53.799 (m - o) und sechs Inserts mit feststehender 3´Position in BAC-Position 54.163 (p - u) generiert, um alle möglichen Kombinationen der *MSY2* Repeats (1 - 4) im Vergleich zu untersuchen. Die Inserts wurden generell in nativer Orientierung in die Vektoren pGL4.10 und pGL4.13, zur Überprüfung auf endogene Promoteraktivität oder Enhancer- und Repressorqualität, kloniert.

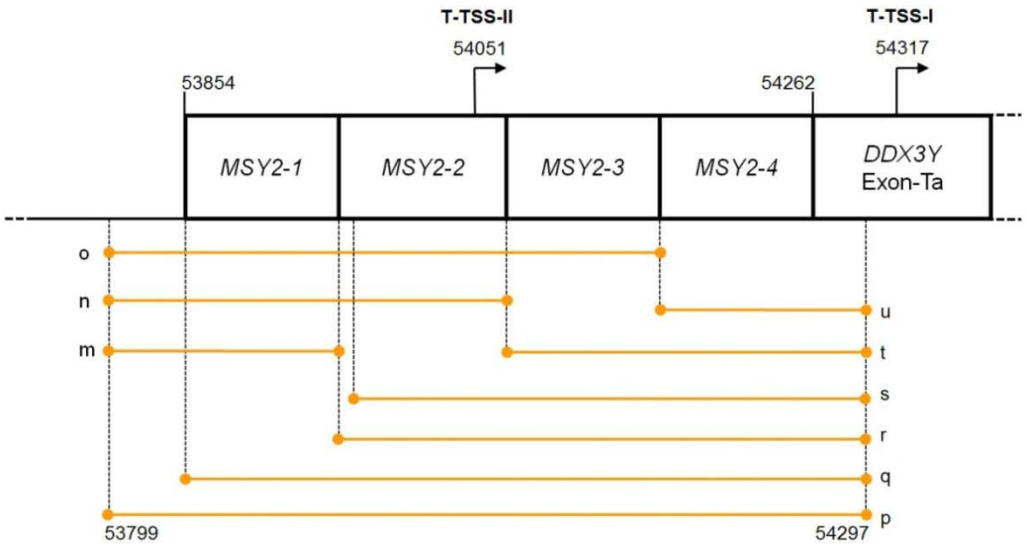


Abb. 4-65: Schematische Darstellung der *DDX3Y MSY2* Konstrukte.
Die in Kapitel 4.9 eingeführte Nummerierung wurde mit den Kleinbuchstaben m bis u für diese DLR-Assay Serie fortgesetzt. Die exakten Positionsbeschreibungen der neun Inserts m bis u sind in Tab. 4-33 aufgeführt. Die Inserts o bis m haben eine fixe 5´ Startposition in BAC-Position 53.799 (BAC-Klon 475I1) und eine variable 3´Länge. Die Inserts p bis u haben eine fixe 3´ Position in BAC-Position 54.297 und variable 5´Längen. Insert s wurde mit demselben Forwardprimer, wie die Inserts t und u, generiert. Die Auswahl der Startposition 53.799 und der End-Position 54.297 außerhalb des *MSY2* Repeats verbesserte die Klonierungsfähigkeit der *MSY2* Repeatregion. Die *DDX3Y* T-TSS-I Position kommt in keinem der Inserts vor.

Mit dem Insert m (*MSY2-1*) im pGL4.13 Vektor konnte eine Aktivitätssteigerung auf 122,4% rel. Luziferaseaktivität gemessen werden (Abb. 4-66, Abb. 4-65). Im promoterlosen Vektor konnte

keine endogene Promoteraktivität für das Insert m festgestellt werden. Das Insert n (*MSY2-1/2*) führte zu einer deutlichen 3,5fachen Reduktion der SV40 Promoteraktivität im pGL4.13 Vektor. Eine endogene Promoteraktivität im pGL4.10 Vektor konnte für Insert n nicht gemessen werden. Mit Insert o (*MSY2-1/2/3*) wurde eine 2,4fache Reduktion der SV40 Promoteraktivität gemessen. Eine deutliche endogene Promoteraktivität ist auch für Insert o nicht meßbar. Von den drei Inserts mit feststehender 5´Position konnte also für alle Inserts eine Modulation der endogenen SV40-Promoteraktivität gezeigt werden. Eine positive Modulation der endogenen SV40-Promoteraktivität war allerdings einzig für Inserts m feststellbar.

Tab. 4-33: Liste der *DDX3Y MSY2* Konstrukte.

Nummerierung	Konstruktbezeichnung	Position im BAC 475I1	Länge des klonierten Fragmentes
m	*MSY2* Repeat 1	53.799 - 53.954	156 bp
n	*MSY2* Repeat 1+2	53.799 - 54.064	266 bp
o	*MSY2* Repeat 1+2+3	53.799 - 54.163	365 bp
p	*MSY2* Repeat 1+2+3+4 (I)	53.799 - 54.297	499 bp
q	*MSY2* Repeat 1+2+3+4 (II)	53.854 - 54.297	444 bp
r	*MSY2* Repeat 2+3+4 (I)	53.955 - 54.297	343 bp
s	*MSY2* Repeat 2+3+4 (II)	53.967 - 54.297	331 bp
t	*MSY2* Repeat 3+4	54.065 - 54.297	233 bp
u	*MSY2* Repeat 4	54.164 - 54.297	134 bp

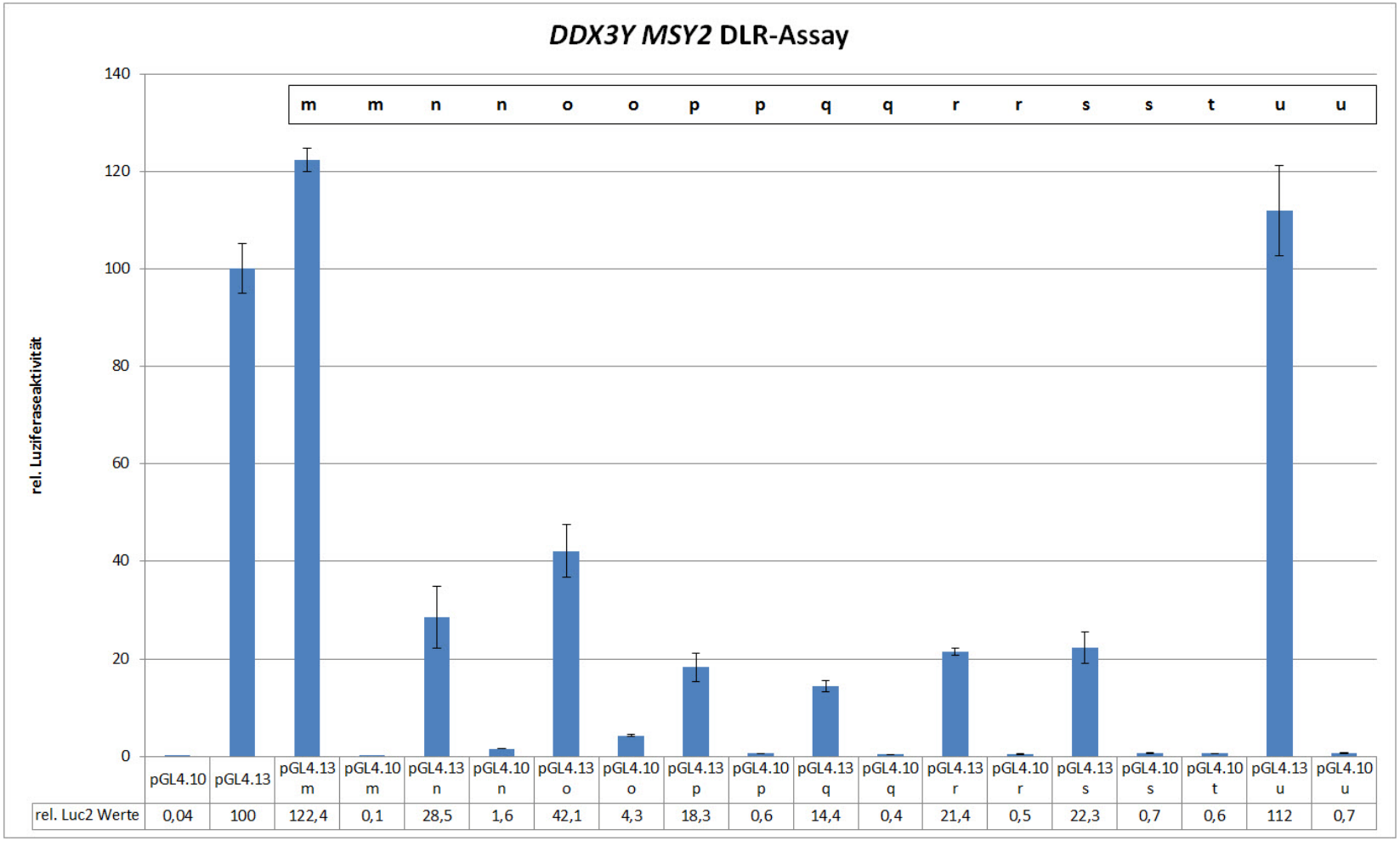
Für das Insert p (*MSY2-1/2/3/4*) mit feststehender 3´Position, das als einziges Insert alle vier Repeateinheiten beinhaltet, wurde ebenfalls eine Aktivitätsreduktion (5,5fach) im pGL4.13 Vektor gemessen. Im promoterlosen pGL4.10 Vektor konnte wiederum keine Eigen-aktivität festgestellt werden (Abb. 4-66, Abb. 4-65). Die Inserts q, r und s ergaben im pGL4.13 Vektor 5,5fache, 6,9fache und 4,7fache Aktivitätsreduktionen (Abb. 4-66). Eine endogene Promoter-aktivität im pGL4.10 Vektor konnte für keines dieser drei Inserts gemessen werden.
Das Insert t (*MSY2-3/4*) konnte auf Grund von Klonierungsschwierigkeiten nicht stabil in nativer Orientierung in den pGL4.13 Vektor eingefügt werden. Im promoterlosen Vektor pGL4.10 konnte keine Promoter-aktivität für Insert t gemessen werden. Für den pGL4.13 Vektor mit Insert u, welches nur die *MSY2* Repeateinheit 4 beinhaltet, konnte eine Steigerung der *Luc*-Promoteraktivität auf 112% im Vergleich zum Kontrollvektor festgestellt werden. Eine Eigenaktivität im pGL4.10 Vektor war auch für Insert u nicht festzustellen.

Insgesamt hat es den Anschein, daß in dieser *DDX3Y MSY2* DLR-Assay Serie nur die zwei Inserts m und u, mit je einer *MSY2*-Repeateinheit, einen positiv regulativen Effekt auf die SV40 Promoteraktivität in der EP2102 Zellinie ausüben. Sobald zwei oder mehr *MSY2* Repeateinheit in den pGL4.13 Vektor inseriert sind, ist ein Repressoreffekt festzustellen. Da

Orientierungseffekte bereits in den Luziferase-Reportergenexperimenten für die *DDX3X* Promotersequenzen festgestellt werden konnten (Abb. 4-63), wurde das *MSY2* Insert p, welches alle vier *MSY2* Repeateinheiten enthält, auch in negativer Orientierung analysiert. Im pGL4.13 Vektor konnte für das Insert p (-Ori) eine rel. Luziferaseaktivität von 0,9%, also eine vollständige Inhibierung der SV40-Promoteraktivität, gemessen werden (nicht gezeigt in Abb. 4-66). Somit ist die Aktivität des pGL4.13 Vektors, mit allen vier *MSY2*-Repeats, deutlich abhängig von der Orientierung der *MSY2*-Repeats.

Tab. 4-34: Absolute und relative Meßwerte der *DDX3Y MSY2* Luziferase-Konstrukte.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
m) pGL4.13 *MSY2* Repeat 1	15.432.352	270.514,3	570.303,7	**122,4**	2,4
m) pGL4.10 *MSY2* Repeat 1	7.353,6	146.358,8	525,5	**0,1**	0,03
n) pGL4.13 *MSY2* Repeat 1+2	3.955.334,3	301.793,7	132.814,6	**28,5**	6,4
n) pGL4.10 *MSY2* Repeat 1+2	246.714	338.988,3	7.226,8	**1,6**	0,07
o) pGL4.13 *MSY2* Repeat 1+2+3	2.490.019,3	126.798,3	196.277	**42,1**	5,4
o)pGL4.10 *MSY2* Repeat 1+2+3	869.983	436.061	20.106,7	**4,3**	0,3
p) pGL4.13 *MSY2* Repeat 1+2+3+4 (I)	1.472.624	173.340,8	85.319,1	**18,3**	2,9
p) pGL4.10 *MSY2* Repeat 1+2+3+4 (I)	48.223,6	186.729,7	2.581,6	**0,6**	0,01
q) pGL4.13 *MSY2* Repeat 1+2+3+4 (II)	1.542.518,7	230.406,7	67.303,7	**14,4**	1,1
q) pGL4.10 *MSY2* Repeat 1+2+3+4 (II)	33.926,8	163.023,8	2.070,1	**0,4**	0,04
r) pGL4.13 *MSY2* Repeat 2+3+4 (I)	1.504.657,8	150.630	99.931,7	**21,4**	0,7
r)pGL4.10 *MSY2* Repeat 2+3+4 (I)	33.151,3	133.551	2.504,1	**0,5**	0,07
s) pGL4.13 *MSY2* Repeat 2+3+4 (II)	1.903.300,5	187.663,5	103.966,1	**22,3**	3,2
s) pGL4.10 *MSY2* Repeat 2+3+4 (II)	42.193	134.711,8	3.064	**0,7**	0,1
t) pGL4.10 *MSY2* Repeat 3+4	45.754,8	177.561,5	2.581,9	**0,6**	0,04
u) pGL4.13 *MSY2* Repeat 4	8.002.715,8	154.546	522.116,7	**112**	9,3
u) pGL4.10 *MSY2* Repeat 4	36.304	131.338,4	3.065,6	**0,7**	0,1



	pGL4.10	pGL4.13	pGL4.13 m	pGL4.10 m	pGL4.13 n	pGL4.10 n	pGL4.13 o	pGL4.10 o	pGL4.13 p	pGL4.10 p	pGL4.13 q	pGL4.10 q	pGL4.13 r	pGL4.10 r	pGL4.13 s	pGL4.10 s	pGL4.10 t	pGL4.13 u	pGL4.10 u
rel. Luc2 Werte	0,04	100	122,4	0,1	28,5	1,6	42,1	4,3	18,3	0,6	14,4	0,4	21,4	0,5	22,3	0,7	0,6	112	0,7

Abb. 4-66: Auswertung des *DDX3Y MSY2* DLR-Assays.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3Y MSY2* Inserts m bis u dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-34 aufgeführt.

4.10.2. Funktionelle Mutationsanalyse des *DDX3Y MSY2* Minisatellitenrepeats

Zwei der, mit Genomatix MatInspector, 128 identifizierten potentiellen TFBSs im *MSY2* Repeat, wurden für eine Mutationsanalyse selektioniert (Abb. 4-67; Tab. 4-35). Die gesamte TFBS-Liste für die 2,5 kb *DDX3Y* Promoterregion ist auf der beiliegenden CD einsehbar. Das für die Auswahl verwendete Programm „Match“, aus dem „TRANSFAC Professional“ Programmpaket (Kel et al., 2003; Matys et al., 2006; Wingender, 2008), nutzt, wie auch das Genomatix-Programm MatInspector, experimentell validierte PVMs („Positional Weight Matrices“). Match operiert ebenfalls mit einem „Matrix-Score“ und einem „TF-Core-Score“ zwischen 0 und 1, wobei 1 wiederum eine perfekte Übereinstimmung zwischen gefundener TFBS und Konsenussequenz angibt. Zusätzlich bietet Match aber optimierte Suchprofile mit vordefinierten Grenzwerten zur Vorhersage potentieller TFBS. Die Auswahl der TFBS, die für eine Mutationsanalyse geeignet erscheinen, erfolgte dann mit dem Profil „Vertebraten, nicht-redundant, Minimum Falsch-Positiv“ (minFP). Dieses Profil vermeidet eine TFBS-Überrepräsentation und gibt daher eher zu wenige als zu viele potentielle TFBS an. Das Profil „minFP“ reduziert somit drastisch die Anzahl an möglichen TFBS und ermöglicht die Auswahl potentiell hochrangiger TFBS, mit der höchsten Bindewahrscheinlichkeit, für die Mutations-analyse. Mit dem Programm „Match“, wurden eine potentielle VDR-RXR TFBS in *DDX3Y MSY2-1* und eine potentielle CREB1 TFBS in *MSY2-2* ausgewählt (Abb. 4-67; Tab. 4-35).

Eine Kurzbeschreibung für das „Vitamin D Response-Element“ (VDE) erfolgte bereits in Kapitel 4.6.3.1. Die hier, in der *MSY2-1* Sequenz, ausgewählte VDR-TFBS ist allerdings nicht mit der X-Y konservierten TFBS aus Kapitel 4.6.3.1 (Tab. 4-10, cTFBS Nr. 13) identisch. Die Mutation wurde jeweils in das Insert m und das Insert n eingeführt. Dabei handelt es sich um dieselbe TFBS-Position, die jeweils in der *MSY2-1* Sequenz beider Inserts lokalisiert ist.
Der zweite selektionierte Transkriptionsfaktor CREB1 („Cyclic AMP-responsive element-binding protein 1”) ist, wie CREM, ein Mitglied der bZIP TF-Familie, der für seine Expression und Funktion in der männlichen Keimbahn bekannt ist (Walker & Habener, 1996; Sassone-Corsi, 1998; De Cesare et al., 1999; Fimia et al., 2000; Don & Stelzer, 2002; Manna et al., 2002). Die selektionierte CREB1-TFBS kommt nicht in der *MSY2-1* Sequenz vor und wurde nur in der *MSY2-2* Sequenz mutiert.
Alle Mutationen wurden mit der PCR-basierten „site-directed Mutagenesis“ Methode eingeführt. In den selektionierten TFBS wurde jeweils ein Nukleotid im TF-Core verändert, so daß die jeweilige Sequenz, durch Match, nicht mehr als TFBS erkannt wurde. Dabei wurde auch darauf geachtet, keine neuen TFBS zu generieren. Zudem wurde bei der Mutationsanalyse darauf geachtet, die Core-Promotermotive in den *MSY2*-Sequenzen nicht zu verändern (Abb. 4-67). Doppelmutationen der VDR- und CREB1-TFBS wurden nicht eingeführt.

Abb. 4-67: Darstellung der eingeführten Mutationen in die *MSY2* und *MSY2-X* Sequenzen.

Eingeführte Mutationen sind durch Pfeil markiert und mit BAC-Position beschrieben. Die Farbgebung in *MSY2-1* und *-2* orientiert sich an Abb. 4-3 (Kapitel 4.1.) und markiert die kartierten Core-Promotermotive. Alle Mutationen wurden mit der PCR-basierten „site-directed Mutagenesis“ Methode eingeführt.

A: VDR-RXR TFBS, BAC-Position 53.913 - 53.937 (25 bp; BAC-Klon 475I1), Match Parameter: „Core-Score“ = 1,0; „Matrix-Score“ = 0,867.

B: CREB1 TFBS, BAC-Position 54.017 - 54.037 (21 bp; BAC-Klon 475I1); Match Parameter: „Core-Score“ = 1,0; „Matrix-Score“ = 0,928.

C: SOX5 TFBS, BAC-Position 29.654 - 29.676 (23 bp; BAC-Klon RP13-13A3); Match Parameter: „Core-Score“ = 1,0; „Matrix-Score“ = 0,982.

D: SMAD3 TFBS, BAC-Position 29.655 - 29.663 (9 bp; BAC-Klon RP13-13A3); Match Parameter: „Core-Score“ = 1,0; „Matrix-Score“ = 1,0.

Tab. 4-35: Liste der *DDX3Y MSY2* Konstrukte mit eingeführter Mutation.
Die in Tab.4-33 eingeführte Nummerierung wurde für die Inserts m und n beibehalten und durch die Abkürzung „Mut“ für Mutation und die betreffende TFBS ergänzt. Die VDR-TFBS ist jeweils nur im *MSY2* Repeat 1 mutiert.

Nummerierung	Konstruktbezeichnung	Position im BAC 475I1	Länge des klonierten Fragmentes
m Mut VDR-RXR	*MSY2* Repeat 1 Mut	53.799 - 53.954	156 bp
n Mut VDR-RXR	*MSY2* Repeat 1+2 Mut VDR	53.799 - 54.064	266 bp
n Mut CREB1	*MSY2* Repeat 1+2 Mut CREB1	53.799 - 54.064	266 bp

In Abb. 4-68 und Tab. 4-36 sind die Ergebnisse der Mutationsanalyse zu sehen. Für das Insert m konnte nach Mutation der VDR-TFBS eine geringe Aktivitätssteigerung auf 131,6% rel. Luziferaseaktivität im pGL4.13 Vektor festgestellt werden, was eine Steigerung von 7,5%, gegenüber dem unmutierten Insert m, bedeutet. Eine eigenständige Promoteraktivität bestand auch nach Mutation nicht im pGL4.10 Vektor.
In Insert n führte die VDR-Mutation ebenfalls zu einer leichten Steigerung von 28% auf nun 36,5% rel. Luziferaseaktivität im pGL4.13 Vektor. Im promoterlosen pGL4.10 ist nach VDR-TFBS Mutation ebenfalls keine Veränderung festzustellen.

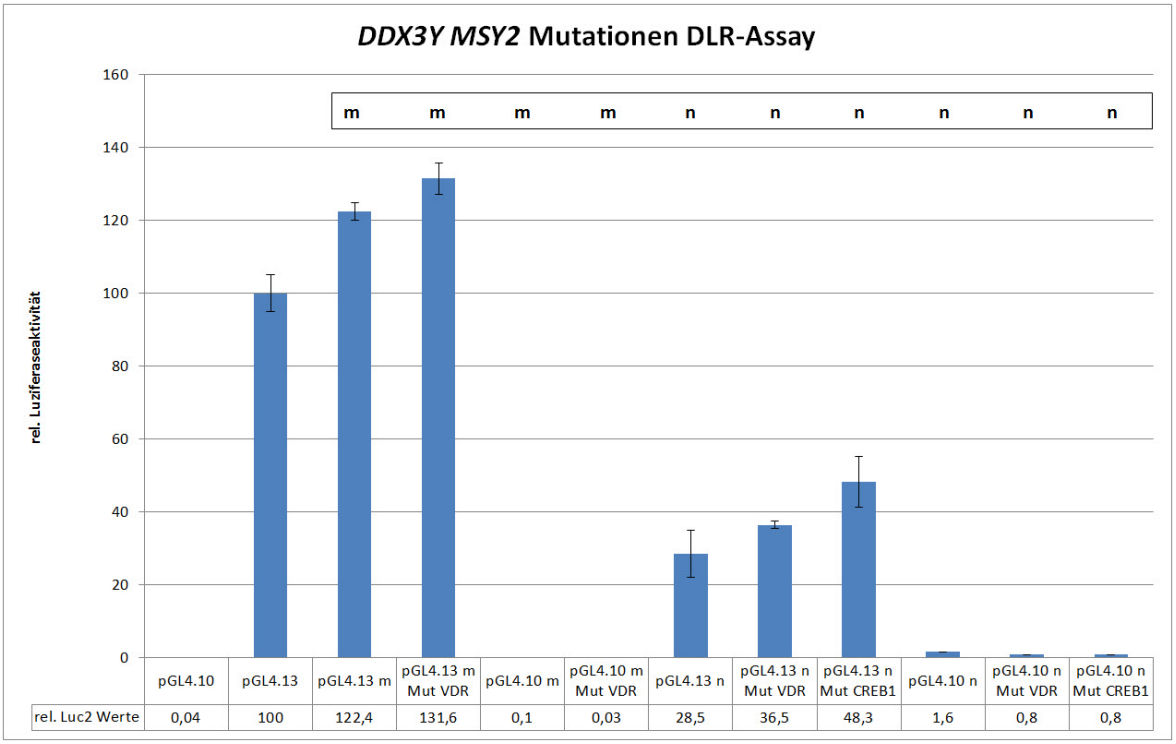


Abb. 4-68: Auswertung der eingeführten Mutationen in die *DDX3Y MSY2* Repeats.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3Y MSY2* Mutationskonstrukt m und n dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-36 aufgeführt.

Für beide Inserts m und n konnte eine Aktivitätssteigerung im pGL4.13 Vektor nach VDR-Mutation nachgewiesen werden. In Insert n bleibt die insgesamt repressive Eigenschaft im

pGL4.13 Vektor aber bestehen. Die VDR-TFBS scheint, in diesem Versuchsaufbau, also eine Repressor-Funktion im nativen *MSY2-1* Repeat zu haben.

Die Mutation der CREB1-TFBS führte in Insert n zu einer 69%igen Aktivitätssteigerung auf 48,3% rel. Luziferaseaktivität im pGL4.13 Vektor. Eine Eigenaktivität im promoterlosen Vektor ist auch für das mutierte Insert n nicht festzustellen. Der beobachtete repressive Charakter des Inserts n, im pGL4.13 Vektor, bleibt auch nach der CREB1-TFBS Mutation bestehen.

Tab. 4-36: Absolute und relative Meßwerte der *DDX3Y MSY2* Mutationskonstrukte.
Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
m) pGL4.13 *MSY2-1* Mut VDR	17.523.520,7	286.843,3	611.846,5	**131,6**	4,3
m) pGL4.10 *MSY2-1* Mut VDR	1.433	95.404,7	150,3	**0,03**	0,001
n) pGL4.13 *MSY2* 1+2 Mut VDR	4.276.873	251.521,7	170.339,1	**36,5**	0,9
n) pGL4.10 *MSY2* 1+2 Mut VDR	225.281	589.860	3.819,5	**0,8**	0,01
n) pGL4.13 *MSY2* 1+2 Mut CREB1	1.939.386	85.326,7	225.505,9	**48,3**	7
n) pGL4.10 *MSY2* 1+2 Mut CREB1	77.593,7	211.935,7	3.598,2	**0,8**	0,05

4.10.3. Funktionelle Analyse der *DDX3X MSY2-X* Sequenz

Auf Grund der Homologie zwischen *MSY2-X* und *MSY2-1* wird auch für die *MSY2-X* Sequenz evaluiert, ob eine regulative Funktion für den *DDX3X* Promoter vorliegt. Die zwei generierten Inserts v und w sind in Abb. 4-69 und Tab. 4-37 angegeben. Die Insertsequenzen wurden so gewählt, daß sie keine bekannten *DDX3X* TSS-Position beinhalten. Die hier generierten Inserts v und w stellen eine nochmalige Verkleinerung des Inserts d (Abb. 4-60, Kapitel 4.9.2.) mit Fokus auf die *MSY2-X* Sequenz dar.

In Abb. 4-70 ist zu sehen, daß mit Insert v, im pGL4.13 Vektor, eine geringe Steigerung der SV40-Promoteraktivität auf 118,1% rel. Luziferaseaktivität erzielt wurde. Eine endogene Promoteraktivität für Insert v ist nicht erkennbar. Im Vergleich zu Insert d (Abb. 4-61, Kapitel 4.9.2.) ist erkennbar, daß der positiv regulative Effekt wesentlich geringer ausfällt. Es ist zu

vermuten, daß entscheidende regulative Elemente in dieser 107 bp langen *MSY2-X* Sequenz, im Vergleich zu Insert d, nicht enthalten sind.

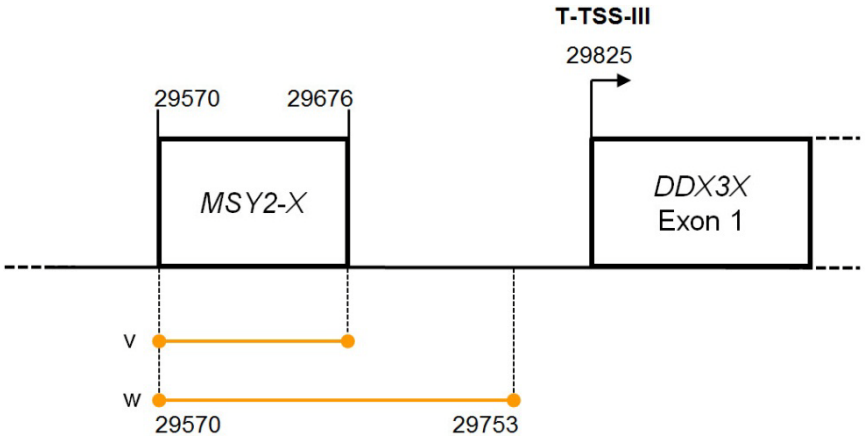


Abb. 4-69: Schematische Darstellung der *DDX3X MSY2-X* Konstrukte.
Insert v enthält exakt die kartierte *MSY2-X* Sequenz. Insert w (*MSY2-X* extended) enthält an der 3´Seite 77 Nukleotide mehr. Das potentielle TF-Modul aus SOX und PAX, das in Kapitel 4.6.3.3. erkannt wurde soll mit dem Insert w untersucht werden. Die *DDX3X* T-TSS-III Position kommt in keinem der Inserts vor.

Mit Insert w, welches generiert wurde, um einen potentiellen Einfluß des kartierten SOX-PAX TF-Moduls an der 3´Grenze der *MSY2-X* Sequenz zu erkennen (siehe Kapitel 4.6.3.3.), konnte eine 2,37fache Aktivitätssteigerung im pGL4.13 Vektor erzielt werden. Im Vergleich zum Insert v, ohne potentielles SOX-PAX TF-Modul, ist dies ein 2facher Anstieg der *Luc*-Promoteraktivität. Für Insert w konnte eine sehr geringe Eigenaktivität von 6,9% rel. Luziferaseaktivität im pGL4.10 Vektor gemessen werden.

Tab. 4-37: Liste der *DDX3X MSY2-X* Konstrukte.
Die in Kapitel 4.10.1 verwendete Nummerierung wurde mit den Kleinbuchstaben v und w für diese DLR-Assay Serie fortgesetzt.

Nummerierung	Konstruktbezeichnung	Position im BAC RP13-13A3	Länge des klonierten Fragmentes
v	*MSY2-X*	29.570 - 29.676	107 bp
w	*MSY2-X* extended	29.570 - 29.753	184 bp

In der verlängerten Insert w Sequenz scheinen also Aktivator-Elemente lokalisiert zu sein, die auch in Insert e (Abb. 4-61, Kapitel 4.9.2.) zu einer 2fachen Aktivitätssteigerung (200,1% rel. Luziferaseaktivität) des SV40-Promoters in der EP2102 Zellinie führten. In der 96 bp langen Schnittmenge (BAC-Klon RP13-13A3, Pos. 29.658 - 29.753) aus Insert w und Insert e sollte, entsprechend dieser Ergebnisse, ein starkes Aktivatorelement der *DDX3X* Promoterdomäne lokalisiert sein. In dieser Region konnten 52 potentielle TFBS (siehe beiliegende CD) und vier TF-Module (Tab. 4-15) kartiert werden, u.a. sieben der 12 potentiellen ZF5-TFBS der 260 bp langen POZ-Region (BAC-Pos. 29.552 bis 29.812). Der Promoterkontext und das zelluläre Umfeld sind für dieses Aktivatorelement wahrscheinlich von funktioneller Bedeutung, da in der

ersten DLR-Assay Serie in Kapitel 4.9.1. (Abb. 4-59) gezeigt werden konnte, daß mit Konstrukt VII, welches in derselben BAC-Position 29.658 wie Insert e beginnt, eine geringere Luziferaseaktivität erzielt werden konnte, als mit dem längeren Promoterkonstrukt VI, das in derselben BAC-Position 29.460, wie das Insert d beginnt.

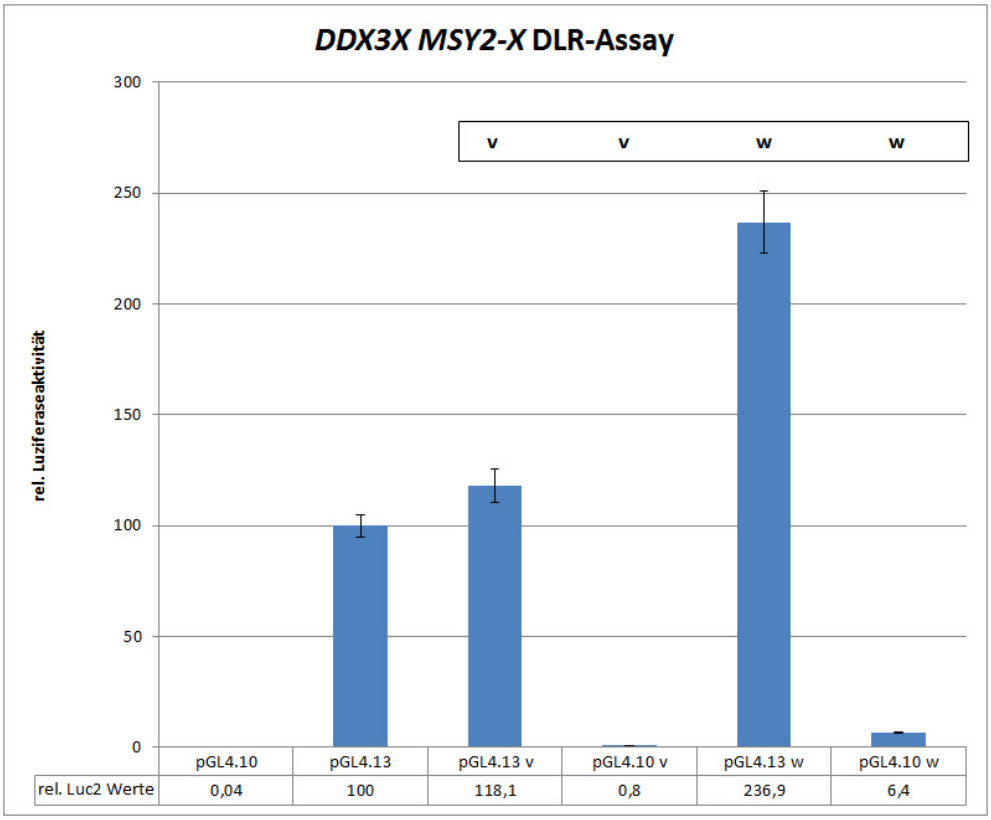


Abb. 4-70: Auswertung des *DDX3X MSY2-X* DLR-Assays.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der *DDX3X MSY2-X* Inserts v und w dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-38 aufgeführt.

Tab. 4-38: Absolute und relative Meßwerte der *DDX3X MSY2-X* Konstrukte.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
v) pGL4.13 *MSY2-X*	16.719.046,7	304.232	550.609,9	**118,1**	7,5
v) pGL4.10 *MSY2-X*	131.478,7	352.603,7	3.742,8	**0,8**	0,05
w) pGL4.13 *MSY2-X* ex.	11.278.064	102.547	1.033.761,2	**236,9**	14
w) pGL4.10 *MSY2-X* ex.	217.203,5	124.839	17.512,6	**3,8**	0,3

4.10.4 Funktionelle Mutationsanalyse der *DDX3X MSY2-X* Sequenz

Analog zur *DDX3Y MSY2-X* Mutationsanalyse wurden für die *DDX3X MSY2-X* Inserts v und w zwei potentielle TFBS mit dem Programm Match zur Mutationsanalyse selektioniert. In den Inserts v und w wurden jeweils die TFBS für SOX5 und SMAD3, durch Einzelnukleotid-austausch, mutiert (siehe Abb. 4-67; Tab. 4-39). Wie für die *DDX3Y MSY2* Inserts wurden auch hier keine Doppelmutanten generiert.

Tab. 4-39: Liste der *DDX3X MSY2-X* Konstrukte mit eingeführter Mutation.
Die in Tab. 4-38 eingeführte Nummerierung wurde für die Inserts v und w beibehalten und durch die Abkürzung „Mut" für Mutation und die betreffende TFBS ergänzt.

Nummerierung	Konstruktbezeichnung	Position im BAC RP13-13A3	Länge des klonierten Fragmentes
v Mut SOX5	*MSY2-X* Mut SOX5	29.570 - 29.676	107 bp
v Mut SMAD3	*MSY2-X* Mut SMAD3	29.570 - 29.676	107 bp
w Mut SOX5	*MSY2-X* Mut ex. SOX5	29.570 - 29.753	184 bp
w Mut SMAD3	*MSY2-X* Mut ex. SMAD3	29.570 - 29.753	184 bp

SOX5 und dessen Funktionalität in der Keimbahn wurden bereits in Kapitel 4.6.3.1. beschrieben. Die hier mutierte SOX5-TFBS ist identisch mit der X-Y konservierten SOX5-TFBS in Kapitel 4.6.3.1. (Tab. 4.10, cTFBS Nr. 10). Das zweite Mutationsziel betrifft die SMAD3-TFBS (Mothers against decapentaplegic homolog 3). Für SMAD3 sind ebenfalls Funktionen in der männlichen Keimbahn bekannt (Luukko et al., 2001; Xu et al., 2003; Itman & Loveland, 2008; Sun et al., 2008).

Die SOX5-TFBS Mutation in Insert v bewirkte einen Aktivitätsanstieg auf 137,7% rel. Luziferaseaktivität im pGL4.13 Vektor, im Vergleich zum unmutierten Insert v, was einer Steigerung von 16,6% entspricht (Abb. 4-71).

Nach SMAD3-TFBS Mutation in Insert v konnte ein Anstieg auf 145,7% rel. Luziferaseaktivität, eine Steigerung von 23,4%, festgestellt werden. Für keine der beiden eingeführten Mutationen konnte eine Veränderung im promoterlosen Vektor gemessen werden.
Sowohl negativ, als auch positiv regulative Effekte konnten bereits für beide TFs SOX5 und SMAD3 beschrieben werden (Gao et al., 2005; Hjelmeland et al., 2005; Martinez-Morales et al., 2010). In dem verlängerten Insert w führte die Mutation der SOX5-TFBS zu einer 7,3%igen Reduktion der rel. Luziferaseaktivität auf 220,7% im pGL4.13 Vektor (Abb. 4-71). Die Mutation der SMAD3-TFBS in Insert w führte zu einer deutlichen Absenkung der rel. Luziferaseaktivität auf 132,8%, was einer Reduktion um 78,4% entspricht. In den promoterlosen Vektoren konnten keine deutlichen Veränderungen registriert werden.

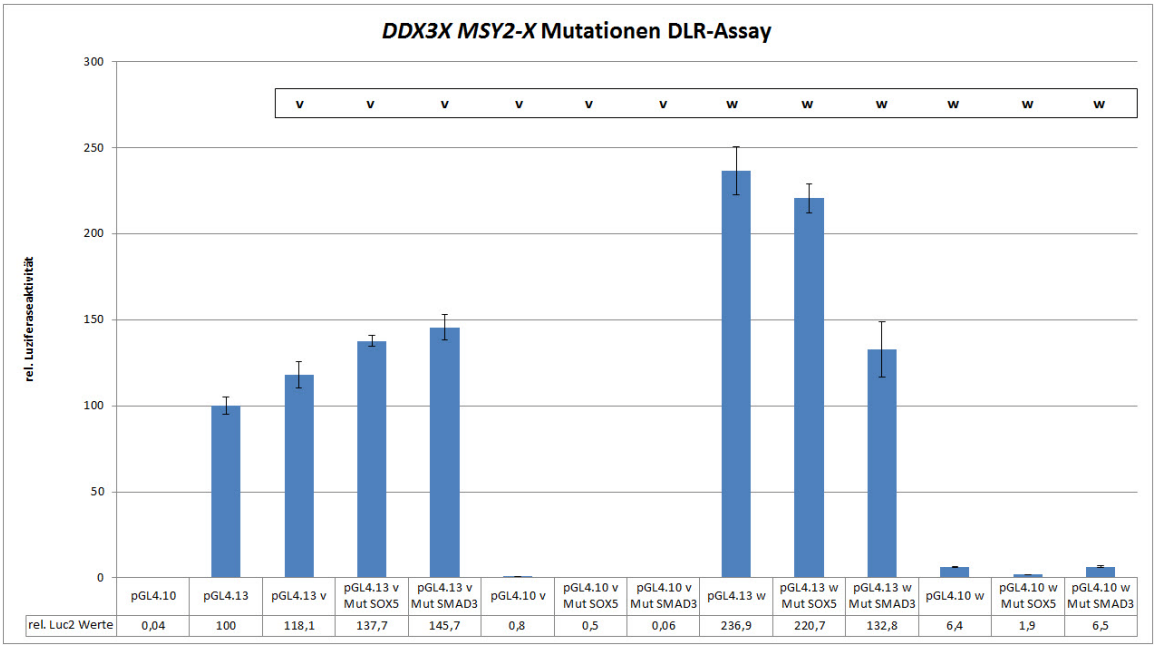


Abb. 4-71: Auswertung der *DDX3X MSY2-X* Mutationskonstrukte.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt. Die experimentellen Werte sind in Bezug zu pGL4.13 dargestellt. In dieser Graphik sind die Ergebnisse der mutierten *DDX3X MSY2-X* Inserts v und w dargestellt. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-40 aufgeführt.

Tab. 4-40: Absolute und relative Meßwerte der *DDX3X MSY2-X* Mutationskonstrukte.

Die Positivkontrolle pGL4.13 ist standardmäßig auf 100% Aktivität eingestellt, so daß alle anderen Werte daran ausgerichtet werden. Die Mittelwerte der absoluten Werte ergeben sich aus mind. drei Einzelmessungen zwei unterschiedlicher Testtage.

Konstrukt	Absolute Firefly Meßwerte im Mittel in LUs	Absolute *Renilla* Meßwerte im Mittel in LUs	Firefly/*Renilla* Quotient in RLUs	Rel. Luziferase-aktivität in %	Standard-abweichung in %
pGL4.10 (neg. Kontr.)	1.076,8	62.282,2	174,3	**0,04**	0,004
pGL4.13 (pos. Kontr.)	2.995.849,4	63.998,8	466.111,9	**100**	5,1
v) pGL4.13 *MSY2-X* Mut SOX5	20.943.044	326.640,7	641.909,8	**137,7**	3,2
v) pGL4.13 *MSY2-X* Mut SMAD3	20.384.581,3	300.940,3	678.911,6	**145,7**	7,6
v) pGL4.10 *MSY2-X* Mut SOX5	83.837	365.900,7	2.290,6	**0,5**	0,002
v) pGL4.10 *MSY2-X* Mut SMAD3	1.289,7	43.138,7	299,1	**0,06**	0,006
w) pGL4.13 *MSY2-X* ex. Mut SOX5	14.761.800	143.605	1.028.898,6	**220,7**	8,4
w) pGL4.13 *MSY2-X* ex. Mut SMAD3	16.202.034	264.694	619.189,4	**132,8**	16,2
w) pGL4.10 *MSY2-X* ex. Mut SOX5	103.198,8	118.988	8677,4	**1,9**	0,2
w) pGL4.10 *MSY2-X* ex. Mut SMAD3	1.278.417,3	426.684,3	30.080,8	**6,5**	0,5

In diesem Mutationsassay fällt auf, daß beide Mutationen in Insert v und Insert w einen gegenteiligen Effekt hervorrufen. In Insert v führten die TFBS-Mutationen zu einer Aktivitäts-

steigerung und in Insert w führten sie jeweils zu einer Reduktion. Mögliche Erklärungen hierfür sind z.B. potentielle TF-Module, an die mehrere TFs kooperativ binden. Je nach angebotenem Interaktionspartner kann die regulative Wirkung positiv oder negativ moduliert werden. Die DNA-Bindung eines TFs ohne Interaktionspartner kann ebenfalls andere Funktionen hervorrufen.

Zusammenfassend läßt sich für die *MSY2* und *MSY2-X* Mutationsanalyse sagen, daß die vier, mit dem Programm Match, selektionierten TFBSs für VDR, CREB1, SOX5 und SMAD3 offensichtlich eine Funktion in der Regulation dieser Sequenzelemente als Aktivator- oder Repressor-TFs ausüben und somit wahrscheinlich auch in den nativen *DDX3Y* und *DDX3X* Promoterdomänen funktionell sind. In wie fern diese Regulation gewebespezifisch realisiert werden kann, läßt sich im Umfeld der EP2102 Zellinie nicht betrachten, aber die Analyse hat gezeigt, daß es sich hierbei tatsächlich um potentiell funktionelle TFBSs handelt.

4.11. Analyse potentieller Translationskontroll-Elemente in der 5´UTR Sequenz der *DDX3X* Transkripte mit möglichem Einfluß auf die Translation in Human-Hodengewebe

Für das Human *DDX3X* Gen konnte experimentell gezeigt werden, daß in Humangeweben, neben einer ubiquitär exprimierten *DDX3X* mRNA mit kurzer 5´UTR (TSS-I), zwei gewebe-spezifische Transkriptklassen (T/B-TSS-II, T-TSS-III) mit verschiedenen 5´UTR Längen existieren. Zur Klärung ob und welchen Einfluß die unterschiedlich langen 5´UTRs der Hoden- und Gehirn-Transkriptvarianten auf die Translationseffizienz der *DDX3X* Transkripte haben, werden die 5´UTRs zunächst *in-silico* auf potentielle Translationskontroll-Elemente und Sekundärstrukturen untersucht. Die *DDX3X* 5´UTR *in-silico* Analyse erfolgt zuerst horizontal im Vergleich zu den 5´UTRs der Human *DDX3Y* Transkripte, um potentielle Unterschiede der *DDX3Y* und *DDX3X* Translationskontrolle erkennen zu können. In einem zweiten Schritt soll evaluiert werden, ob die ermittelten Kontrollelemente humanspezifisch, oder evolutionär konserviert sind. Dazu werden die *DDX3X* 5´UTRs *in-silico* mit den homologen Transkripten von Schimpanse (*Pan troglodytes* = Ptro), Gorilla (*Gorilla gorilla* = Ggor), Orang-Utan (*Pongo pygmaeus* = Ppyg), Rhesusaffe (*Macaca mulatta* = Mmul), Weißbüschelaffe (*Callithrix jacchus* = Cjac) und Mausmaki (*Microcebus murinus* = Mmur), sowie Rind (*Bos taurus* = Btau) und Maus (*Mus musculus* = Mmus) verglichen. Vom Mausmaki waren X-chromosomale Sequenzen erhältlich, so daß auch ein Vertreter der evolutionär entfernteren Primaten-Unterordnung Strepsirrhini in die *in-silico* Analyse einbezogen werden konnte. Die funktionelle Charakterisierung der *DDX3X* 5´UTR-Varianten erfolgt anschließend in einem *in-vitro* Translations-Assay.

4.11.1. *DDX3X* 5´UTR Sequenzelemente mit potentieller Funktion für die Translationskontrolle in Hodengewebe

In der hodenspezifischen Human *DDX3X* T-TSS-III 5´UTR konnten unterschiedliche *cis*-Faktoren sowie strukturelle 5´UTR-Elemente festgestellt werden. Diese potentiell translations-regulativen Elemente werden im Folgenden im Vergleich zu den *DDX3Y* 5´UTR-Varianten dargestellt. Der in Abb. 4-72 gezeigte ClustalW-Sequenzvergleich basiert auf dem Alignment in Kapitel 4.3. (Abb. 4-14). Als ein translationsregulierender Faktor kommt die 5´UTR-Länge in Frage. Per se ist die Länge einer 5´UTR nicht zwingend ein negativer Faktor, aber je länger eine Leadersequenz ist, desto mehr und stabilere Sekundärstrukturen können auftreten, die das Ribosomen-Scannen erschweren oder unterbinden können und somit die Translationseffizienz herabsetzen können (Kozak, 2005).

Die Durchschnittslänge einer humanen 5´UTR beträgt 170 Nukleotide (Calvo et al., 2009). Die *DDX3X* 5´UTR liegt mit max. 905 Nukleotiden also weit über diesem Wert. Die längste mögliche *DDX3Y* 5´UTR, ausgehend von T-TSS-II an BAC-Position 54.051 und mit Tb-1b Splicing, erreicht eine Länge von 674 Nukleotiden.
Ein deutlicher Unterschied in den *DDX3X* und *DDX3Y* 5´UTR-Domänen ist auch in der Anzahl der jeweiligen uORFs zu finden. Für das *DDX3X* T-TSS-III Transkript sind in der durchgehenden Variante zwei uORFs, in der gesplicten Variante (Kapitel 4.2.3, Abb. 4-13) ist ein uORF identifizierbar (Abb. 4-72).Der *DDX3X* uORF1 ist 78 bp lang (Pos. 30.004 – 30.081) und kodiert für ein Peptid von 25 AS. Der *DDX3X* uORF2, der nur in der ungesplicten *DDX3X* 5´UTR-Sequenz vorkommt, ist 36 bp lang (Pos. 30257 – 30292) und kodiert für ein Peptid von 11 AS. Beide uORFs stehen nicht mit dem DDX3X Protein-Leserahmen in Verbindung und führen nicht zu DDX3X Proteinvarianten. In den *DDX3X* T/B-TSS-II und TSS-I Transkripten sind keine uORFs vorhanden.
In *DDX3Y* T-TSS-II Transkripten sind insgesamt fünf uORFs kartierbar (Jaroszynski et al., 2011). Jeweils ein uORF befindet sich in der *MSY2*-Repeateinheit 3 und 4 (je 24 AS), zwei befinden sich stromaufwärts der T-TSS-I Initiationsstelle (13 und 18 AS) und ein uORF ist am 3´Ende von Exon-Ta (23 AS) lokalisiert. Zusätzlich ist in *MSY2-3* und *MSY2-4* je ein uATG vorhanden, welches unmittelbar von einem TAA Stopcodon gefolgt wird. In *DDX3Y* T-TSS-I Transkripten ist ein uORF (uORF Nr. 5) direkt vor dem Exon-Ta 5´Splicesignal vorhanden. Die uORFs stehen nicht mit dem DDX3Y Protein-Leserahmen in Verbindung und führen nicht zu DDX3Y Proteinvarianten. In den *DDX3Y* TSS-I-(T) und TSS-I Transkripten sind keine uORFs vorhanden. In den *DDX3X* und *DDX3Y* 5´UTR-Sequenzen ist nur ein uORF (*DDX3X* uORF1; *DDX3Y* uORF5) konserviert (Abb. 4-72).

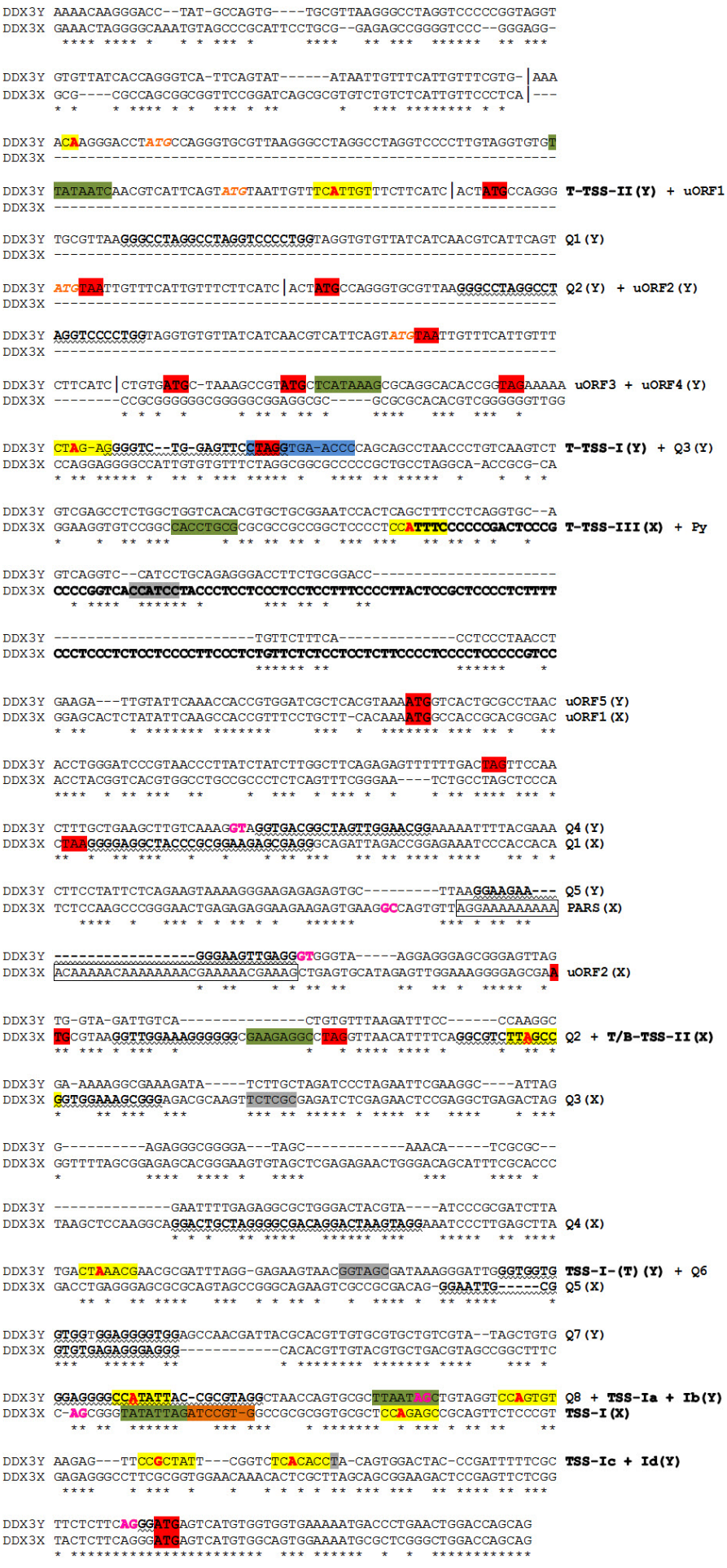


Abb. 4-72: Vergleich der genomischen *DDX3Y* und *DDX3X* Sequenzen in den 5´UTR-Regionen.

ClustalW-Alignment der *DDX3Y* und *DDX3X* 5´UTR-Regionen. *DDX3Y* Sequenz von 53.854 bis 55.190 (1.337 bp; BAC 475I1) und *DDX3X* Sequenz von 29.570-30.774 (1.205 bp; BAC RP13-13A3). Kartierte *DDX3Y* und *DDX3X* Inr-Motive sind gelb markiert, TSSs mit roten Buchstaben A und G hervorgehoben. Pot. CA-Dinukleotid in *MSY2-2* ist gelb markiert. TATA-Box Motive sind grün, DPEs sind hellgrau, BRE[d] ist orangefarben und MTE blau markiert. Start- und Stopcodons sind rot hervorgehoben. uATGs sind kursiv und orange dargestellt. 5´- und 3´-Splicesignale sind in Fettschrift und pink angegeben. Die *DDX3X*-spezifische Pyrimidin-reiche Region (Py) ist in Fettschrift, die *DDX3X*-spezifische Adenin-reiche Sequenz (PARS) mit einem Kasten hervorgehoben. Potentielle QGRSs (Q) sind durchnummeriert und in Fettschrift und gewellt unterstrichen dargestellt. Die vier Y-spezifischen *MSY2* Repeats sind durch vertikale Striche voneinander getrennt. Weitere Erklärungen siehe Text.

Ein Vergleich der für die Nutzung der ORFs (uORF1, uORF2, *DDX3X* ORF) notwendigen Kozak-Sequenzen (ATG-Kontext = gcc**R**ccATG**G**) verdeutlicht, daß uORF1 einen optimalen Translationskontext mit der Sequenz **A**AAATG**G** (-3 = A; +4 = G) verfügt. Für uORF2 wurde nur ein schwaches ATG-Initiationspotential (**C**GAATG**C**) festgestellt. In den zwei kritischen Positionen (-3 = C; +4 = C) kommt nicht die optimale Konsensussequenz vor. Der Kontext des Haupt-Leserahmens (**G**GGATG**A**) ist dagegen adäquat, d.h. es stimmt eine der beiden kritischen Positionen (-3 = G; +4 = A) mit der Konsensussequenz überein. Es besteht somit die Möglichkeit, daß insbesondere uORF1 Einfluß auf die Translationseffizienz des nachfolgenden *DDX3X* Haupt-ATGs in der langen *DDX3X* 5´UTR-Variante nehmen kann. In diesem Fall würde die „First ATG"-Regel zum Tragen kommen, die besagt, daß das Ribosom, beim Scannen entlang der 5´UTR, die Translation am ersten ATG mit starkem Kontext initiiert.
Die *DDX3Y* uORFs 1 bis 4 verfügen über einen adäquaten ATG-Kontext. Das *DDX3Y* uORF5 verfügt, wie der positionshomologe *DDX3X* uORF1, über einen starken Translationskontext. Der Initiationskontext des *DDX3Y* Haupt-Leserahmens ist identisch mit *DDX3*X, also adäquat.

Das *DDX3X* uORF2 ist vermutlich durch eine 26 nts lange Duplikation in der genomischen Sequenz entstanden (Abb. 4-73). Die Sequenz in der *DDX3X* 5´UTR zwischen BAC-Position 30.236 und 30.261 scheint zwischen BAC-Position 30.262 und 30.287 dupliziert worden zu sein. 20 der 26 Nukleotide sind identisch in der duplizierten Sequenz wiederzufinden.

Human BAC RP13-13A3, *DDX3X* 5´UTR

```
      Pos. 30236        uORF2  Pos. 30261
GTGCA|TAGAGTTGGAAAGGGGAGCGAATGCG
      TAAGGTTGGAAAGGGGGGCGAAGAGG|CCTAG
      **  ********** *****   *
      Pos. 30262               Pos. 30287
```

Abb. 4-73: Entstehung von *DDX3X* uORF2 durch potentielle Duplikation in der Human *DDX3X* 5´UTR Sequenz.
Die Duplikation ist als Sequenzalignment gezeigt, um homologe Positionen darzustellen. Sterne unterhalb des Alignments zeigen homologe Nukleotide an. Translationsstart- und Stopcodon sind rot markiert. Die duplizierte Sequenz ist durch horizontale Striche eingegrenzt. Positionsbezeichnungen beziehen sich auf BAC RP13-13A3.

Die Lage der *DDX3X* uORF1 bzw. *DDX3Y* uORF5 Kozak-Sequenzen stimmt mit der Position der X-Y konservierten YY1-TFBS überein (siehe Kapitel 4.6.3.1, Tab. 4-10). Die YY1 TF-Konsensussequenz „C[AAGATGGC]GGC" stimmt mit der Kozak-Sequenz in den kritischen Positionen -3, ATG und +4 (unterstrichen) überein (Xi et al., 2007). Ebenfalls wurde von Elfakess & Dikstein (2008) das Translationsinitiationselement „TISU" (Translation Initiator of Short 5´UTR) mit der Sequenz „AAGATGGC" beschrieben, was zu 100% mit der YY1 TF-

Bindesequenz (Homologie durch eckige Klammer in YY1-TFBS gezeigt) und weniger mit der Kozak-Sequenz für optimale Translationsinitiation übereinstimmt.

Tab. 4-41: Vergleich von *DDX3X* uORF1 und *DDX3Y* uORF5 ATG-Kontext zu Kozak-Sequenz, TISU-Element und YY1 TF-Bindesequenz.

Das Triplet „ATG" ist jeweils unterstrichen, die kritischen Positionen -3 und +4 für die Translationsinitiation sind in Fettdruck hervorgehoben. Sequenzhomologie in allen fünf *cis*-Elementen ist unterhalb der Sequenzen durch Sterne markiert.

***cis*-Element**	**Sequenz**
DDX3X uORF1	CAC**A**AAATG**G**
DDX3Y uORF5	CGT**A**AAATG**G**
Kozak-Sequenz	gcc**R**ccATG**G**
TISU-Element	**A**AGATG**G**C
YY1 TF-Konsensusbindestelle	C**A**AGATG**G**CGGC * ****

In Tab. 4-41 werden die fünf genannten Elemente verglichen. Es wird deutlich, daß in den entscheidenden Positionen (-3, ATG, +4) aller fünf Sequenzen eine 100%ige Übereinstimmung besteht. In wie fern das TISU-Element auch eine Funktion in langen 5´UTRs haben kann ist bisher nicht bekannt.

Als weiteres putatives Translations-Kontrollelement ist die 137 bp lange *DDX3X*-spezifische Pyrimidin-reiche Sequenzdomäne (BAC-Pos. 29.826 – 29.962) am distalen 5´Ende der *DDX3X* T-TSS-III 5´UTR-Variante zu nennen. In der homologen *DDX3Y* 5´ Region ist keine Pyrimidin-Sequenz vorhanden (Kapitel 4.6.2).

mRNAs mit 5´terminalen Pyrimidin-reichen Sequenzen gehören potentiell zu den 5´TOP-Transkripten („Terminal Oligopyrimidin Tract"; Avni et al., 1997; Hamilton et al., 2006). 5´TOPs werden häufig in mRNAs gefunden, die für Proteine kodieren, welche nachfolgend in der Translationsmaschinerie eine wichtige Rolle spielen (z. B. ribosomale Proteine, Elongationsfaktoren, Poly(A)-bindende Proteine). Diese Proteine unterliegen selbst einer strengen Expressionskontrolle auf Ebene der Translation. Diese Art der translationalen Kontrolle ist, z.B. als Folge von Aminosäuremangel, gekennzeichnet durch einen selektiven "Shift" der 5´TOP-mRNAs aus translationsaktiven Fraktionen (Polysomenfraktionen), hin zu inaktiven mRNP-gebundenen mRNA-Formen (subpolysomale Fraktionen; Hamilton et al., 2006). Eine funktionelle 5´TOP-Sequenz muß drei Kriterien erfüllen und wie folgt aufgebaut sein: An die 5´Cap-Struktur (m^7Gppp) schließt als erstes Nukleotid ein Cytosin an (i), welches unmittelbar von 4 bis 14 (ii) ununterbrochenen (iii) Pyrimidinresten gefolgt wird (Hamilton et al., 2006; Meyuhas, 2000).

Die Analyse, auf diese 5´TOP-Signatur, der *DDX3X* 5´UTR-Sequenz mit Start in Position T-TSS-III (BAC-Pos. 29.825), zeigte, daß nicht alle drei Kriterien erfüllt werden. Diese Transkripte beinhalten am distalen 5´Ende bp zwar die 137 bp lange Pyrimidin-reiche Sequenz (Abb. 4-72), beginnen aber mit einem Adeninrest (gefolgt von neun Pyrimidinen) und nicht mit einem Cytosin.

Weiterhin fällt in der *DDX3X* 5´UTR-Sequenz eine fast vollständig homopolymere Adenin-Sequenz [$A(G)_2(A)_{10}C(A)_5C(A)_8CG(A)_5CG(A)_3G$] von 41 Nukleotiden auf (eingerahmt in Abb. 4-72). Diese „Polypurin (A)-reiche Sequenz“ (PARS) erstreckt sich von BAC-Position 30.186 bis 30.226, zwischen dem 5´Splice-Signal und dem uORF2. Die PARS kommt folglich nur in dem ungesplicten *DDX3X* T-TSS-III Transkript vor und ist *DDX3X* spezifisch. In der *DDX3Y* 5´UTR-Sequenz kommt allerdings, direkt vor dem Exon-Tb Splicesignal, ein kürzerer, 15 Nukleotide langer, R-Trakt positionshomolog vor (Abb. 4-72).

Wie in Kapitel 4.6.2 für die Transkription beschrieben, können binäre und homopolymere Nukleotidkompositionen auch in 5´UTR-Sequenzen Einfluß, auf die Translationsfähigkeit der Transkripte, haben. In Tab. 4-42 sind alle binären Sequenz-Trakte zusammengestellt, die in den unterschiedlichen *DDX3X* und *DDX3Y* 5´UTR-Varianten vorkommen können.

Tab. 4-42: Vergleich binärer „Sequenz-Trakte“ zwischen den *DDX3X* und *DDX3Y* 5´UTR-Varianten.
Sequenz-Trakte werden entsprechend der jeweiligen BAC-Position und des IUPAC-Codes aufgeführt. Es werden nur binäre Sequenz-Trakte mit ununterbrochenen Sequenzen von 10 und mehr Nukleotiden (ohne Mismatch) aufgeführt.

DDX3X			***DDX3Y***		
Trakt	**Position & Transkript**	**Länge**	**Trakt**	**Position & Transkript**	**Länge**
S	29.839 (T-TSS-III)	10	**W**	54.648 (Ex. Tb)	10
Y	29.860 (T-TSS-III)	25	**R**	54.681 (Ex. Tb)	16
Y	29.891 (T-TSS-III)	37	**R**	54.702 (Ex. Tb)	15
Y	29.929 (T-TSS-III)	30	**K**	54.955 (TSS-I-(T))	17
R	30.156 (T-TSS-III)	16			
R	30.186 (T-TSS-III)	13			
M	30.189 (T-TSS-III)	26			
W (nur A)	30.189 (T-TSS-III)	10			
R	30.243 (T-TSS-III)	11			
R	30.269 (T-TSS-III)	11			
R	30.556 (T-TSS-III; T/B-TSS-II)	11			
S	30.621 (T-TSS-III; T/B-TSS-II)	10			

Für *DDX3X* können zwei Regionen mit unterschiedlichem Charakter definiert werden. Der distale 5´UTR-Bereich ist vor allem (C,T)-reich, der zweite Sequenzbereich, um das 5´Splicesignal und das *DDX3X* uORF2, ist dagegen vor allem (A,G)-reich. Zehn von 12 kartierten binären *DDX3X* 5´UTR-Sequenzabschnitten befinden sich stromaufwärts der T/B-TSS-II Initiationsstelle. Für die *DDX3Y* 5´UTR ist auffallend, daß binäre Sequenzen nur in Exon-Tb bzw. im Bereich von TSS-I-(T) vorkommen, eben die Bereiche, die im Haupt-Spliceprodukt Exon-Ta-1b nicht vorhanden sind. Der Vergleich zeigt also, daß binäre Sequenzen in den *DDX3Y* 5´UTR-Varianten in wesentlich geringerem Maße, als in der *DDX3X* 5´UTR vorkommen.

4.11.2. Strukturelle *cis*-Elemente in der *DDX3X* und *DDX3Y* 5´UTR-Sequenz mit potentieller Translationskontrollfunktion

In der Basenzusammensetzung und dem (C,G)-Gehalt unterscheiden sich die 5´UTRs von *DDX3X* und *DDX3Y* deutlich (Tab. 4-44). Die verschiedenen *DDX3X* 5´UTR-Längen weisen alle einen (C,G)-Gehalt zwischen 57,7 und 61% auf, wohingegen für *DDX3Y* (C,G)-Werte von 48,07% bis 52,8% zu finden sind. Dieser Unterschied wird vermutlich Einfluß auf die Sekundärstruktur der Transkripte und somit auf deren Translationsfähigkeit haben. Nukleinsäuren gelten ab einem (C,G)-Gehalt von 50% als (C,G)-reich (Bajic et al., 2006), was wiederum zu einer vermehrten Bildung von Sekundärstrukturen führen kann.
G-Quadruplex-Strukturen, die in „Guanin-reichen Sequenzen“ (GRS) entstehen können, sind als strukturell regulative *cis*-Elemente in mRNAs bekannt (Wieland & Hartig, 2007; Neidle, 2009). Sie bilden eine lokal begrenzte viersträngige Helix aus, die über die Bildung von Hoogsteen-Wasserstoffbrücken stabilisiert werden (Munk, 2010). G-Quadruplexe in DNA und mRNA stehen in Zusammenhang mit „Transcription Pausing“ (Yonaha & Proudfoot, 1999), mRNA Stabilität (Simonsson, 2001), FMRP-Bindung an mRNAs (Darnell et al., 2001), Translationsinitiation durch mögliche IRES-Strukturen (Bonnal et al., 2003; Morris et al., 2010), sowie der Translationsrepression durch Unterbinden des Scanning-Mechanismuses (Oliver et al., 2000). Obwohl in einigen Publikationen nur von G-Quadruplexen gesprochen wird, wenn die G-Tetraden-Gruppen aus jeweils drei Guaninen (G_3N) bestehen, beschreiben aktuelle Publikationen auch stabile G-Quadruplex Strukturen mit zwei Guaninen in G_2N-Konformation (Lim et al., 2009; Kankia et al., 2005; Kikin et al., 2006; Zarudnaya et al., 2003).

Für den Vergleich der *DDX3X* und *DDX3Y* 5´UTRs auf potentielle „Quadruplex forming G-rich Sequences“ (QGRS), auch „putative guadruplex sequence (PQS) genannt, wurde das Programm „QGRS-Mapper“ verwendet (Kikin et al., 2006). „QGRS-Mapper“ erkennt potentielle QGRS mit dem Motiv „$G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$“, wobei x für die G-Tetraden und y für die Loop-

Längen stehen. Für die Berechnung der QGRS-Stabilität verwendet QGRS-Mapper also zwei Kriterien, die dann als „G-Score“, einem Wahrscheinlichkeitswert für die Ausbildung stabiler G-Quadruplexe, in Abhängigkeit der maximalen QGRS-Länge, angegeben werden. Zum einen wird die Größe der G-Gruppen betrachtet (2 bis 6 Guaninreste in einem Quartett), wobei die größeren G-Gruppen stabiler sind. Zum zweiten wird die Länge der dazwischen liegenden Sequenzen („Loops“) berücksichtigt. Hierbei gilt, daß stabile QGRS möglichst gleichlange Loops beinhalten sollen und daß größere G-Tetraden höhere G-Scores erzielen (Kikin et al., 2006).

In Tab. 4-43 sind die fünf nicht überlappenden *DDX3X* und die acht nicht überlappenden *DDX3Y* QGRS aufgelistet, die durch QGRS-Mapper in den Human *DDX3X* und *DDX3Y* 5´UTR-Sequenzen vorhergesagt werden. G-Scores für stabile QGRS mit G_2N Struktur haben einen Wert von 20 bzw. 21 (grau unterlegte Felder in Tab. 4-43; persönliche Mitteilung L. D´Antonio, 2009).

Tab. 4-43: Kartierung potentieller G-Quadruplexsequenzen in den Human *DDX3X*/*DDX3Y* 5´UTRs.
QGRS mit potentiell stabilen G-Quadruplex Strukturen sind grau unterlegt. „QGRS-Mapper“ Parameter: Max. QGRS-Länge = 30 Nukleotide, Min. G-Gruppen x = 2, Loop-Größe y = 0 bis 36 Nukleotide, ein Loop darf y = 0 werden, nicht überlappende QGRS. Die G-Gruppen sind in den jeweiligen Sequenzen unterstrichen. Der theoretisch maximale G-Score, mit den gewählten Standardparametern, beträgt 105 für eine Sequenz mit x = 6 und y1, y2, y3 = 1.

Nr.		BAC-Position	QGRS Länge	QGRS Sequenz	G-Score	Entfernung von 5´Cap	TSS & 5´UTR-Variante
DDX3X	1	30.082	27	GGGGAGGCTACCCGCGGAAGAGCGAGG	16	258 bp	T-TSS-III gesplict + ungesplict
	2	30.265	15	GGTTGGAAAGGGGGG	20	440 bp	T-TSS-III ungesplict
	3	30.306	25	GGCGTCTTAGCCGGTGGAAAGCGGG	12	482 bp	T-TSS-III ungesplict
	4	30.452	30	GGACTGCTAGGGGCGACAGGACTAAGTAGG	20	628 bp, 139 bp	T-TSS-III ungesplict, T/B-TSS-II
	5	30.543	24	GGAATTGCGGTGTGAGAGGGAGGG	17	719 bp, 230 bp	T-TSS-III ungesplict, T/B-TSS-II
DDX3Y	1	54.085	23	GGGCCTAGGCCTAGGTCCCCTGG	19	35 bp	T-TSS-II
	2	54.184	23	GGGCCTAGGCCTAGGTCCCCTGG	19	134 bp	T-TSS-II
	3	54.320	19	GGGGTCTGGAGTTCCTAGG	13	270 bp, 4 bp	T-TSS-II, T-TSS-I
	4	54.627	21	GGTGACGGCTAGTTGGAACGG	18	577 bp, 311 bp	T-TSS-II, T-TSS-I (Ex. Tb-1a)
	5	54.704	21	GGAAGAAGGGAAGTTGAGGGG	14	654 bp, 388 bp	T-TSS-II, T-TSS-I (Ex. Tb-1a)
	6	54.958	11	GGTGGTGGTGG	21	48 bp	TSS-I-(T)
	7	54.970	10	GGAGGGGTGG	20	60 bp	TSS-I-(T)
	8	55.023	24	GGAGGGGCCATATTACCGCGTAGG	6	113 bp	TSS-I-(T)

Für *DDX3X* wurden drei der fünf QGRS stromaufwärts der T/B-TSS-II Transkriptionsstartstelle kartiert, also in dem 5´UTR-Abschnitt, der hodenspezifisch exprimiert wird. Zudem kommen die *DDX3X* QGRS2 - 5 nur in dem ungesplicten T-TSS-III Transkript vor. Im ubiquitären *DDX3X* TSS-I Transkript ist kein QGRS lokalisiert.
Im *DDX3Y* TSS-I-(T) Transkript fällt ein nicht-kodierender $(TGG)_4$ Trinukleotidrepeat (TNR) auf, der in einer (G,T)-reichen Sequenz „GGGATTGGGTGGTGGTGGTGGAGGGGTGGAG“ (31 bp; BAC-Position 54951 - 54981) lokalisiert ist (siehe auch Kapitel 4.1.1 u. 4.6.2). Der vorliegende $(TGG)_4$-Repeat (unterstrichen in Sequenz oben) im *DDX3Y* TSS-I-(T) Transkript kann auch als QGRS interpretiert werden. Im Bereich des $(TGG)_4$-Repeats konnten zwei potentielle QGRS identifiziert werden. Die Sequenzen „GGTGGTGGTGG“ (G-Score = 21; QGRS6 in Tab. 4-43) und „GGAGGGGTGG“ (G-Score = 20; QGRS7) können potentiell stabile Quadruplex-Strukturen des Typs G_2N bilden. In *DDX3X* Transkripten mit langer 5´UTR und Start in T-TSS-III oder T/B-TSS-II, ist an vergleichbarer Position ebenfalls ein QGRS (*DDX3X* QGRS5) vorausgesagt. Die Sequenzen im Bereich der *DDX3Y* QGRS6 und -7 sind nicht komplett homolog zu *DDX3X* QGRS5, das Potential zur G-Quadruplexbildung ist aber in beiden Sequenzen konserviert (Abb. 4-72).

In vergleichbarer Position zu *DDX3Y* QGRS4 ist ebenfalls ein QGRS in der *DDX3X* 5´UTR-Sequenz (*DDX3X* QGRS1) vorhergesagt. Auch in diesem Fall ist nur die Lage vergleichbar und das Potential zur G-Quadruplexbildung. Es könnte sich dabei also um ein konserviertes Strukturelement oder um eine Konvergenz in der regulativen Funktion handeln.
Weitere *DDX3Y* G-Quadruplex Strukturen sind in den *MSY2*-Repeats 3 und 4, und unmittelbar nach der *DDX3Y* T-TSS-II Position, sowie unmittelbar vor dem ATG Startcodon (QGRS5 nur in der Exon1b Splice-Variante) vorhergesagt. Für beide Humangene sind QGRS somit nur für die gewebespezifischen 5´UTR-Verlängerungen der spezifischen Transkriptvarianten vorhergesagt, nicht aber für die ubiquitär exprimierten Transkripte. Es besteht also die Möglichkeit, daß es sich hierbei um Translationskontrollelemente der gewebespezifischen *DDX3X* und *DDX3Y* Transkripte handelt.

Die Entfernung eines QGRS zur 5´Cap-Struktur einer mRNA gilt als kritisch für die Bildung des Translationsinitiationskomplexes (Huppert et al., 2008). Statistisch kommen QGRS in den ersten 50 bp einer 5´UTR, im Vergleich zum Rest der 5´UTR, gehäuft vor und können die Bildung des 43S-Ribosoms an der 5´Cap-Struktur besonders effektiv unterbinden (Huppert et al., 2008). Für die *DDX3Y* Transkriptvarianten konnten drei 5´UTRs identifiziert werden, in denen QGRS innerhalb dieses 50 bp Fensters lokalisiert sind (Tab. 4-43, gelb unterlegt). Dies sind Transkripte mit Start in *DDX3Y* T-TSS-II, T-TSS-I und TSS-I-(T), also alle *DDX3Y* Transkriptvarianten, die ausschließlich oder prädominant in Keimzellen exprimiert werden.

In *DDX3X* konnte kein QGRS innerhalb des kritischen 50 bp Fensters lokalisiert werden, was aber nicht bedeutet, daß stromabwärts gelegene QGRS den Scanning-Mechanismus ebenfalls effektiv unterbrechen können.

Da in der *DDX3Y* TSS-I-(T) 5´UTR-Variante bereits der Trinukleotidrepeat (TNR) auffiel, wurden die weiteren *DDX3Y* und *DDX3X* 5´UTR-Varianten ebenfalls auf Anwesenheit kurzer repetitiver Sequenzen untersucht. Es ist bekannt, daß die genomische Instabilität expandierender TNRs für eine Reihe neurodegenerativer Erkrankungen wie „Chorea Huntington“ und „Fragiles X-Syndrom“ verantwortlich sind (Brouwer et al., 2009; McMurray, 2010). Lawson & Zhang (2008) beschreiben solche „Simple Sequence Repeats“ (SSRs), auch Mikrosatelliten genannt, als generell Expressions-regulative *cis*-Elemente, da SSRs signifikant häufiger in 5´UTRs, als in anderen Genbereichen, auftreten. Innerhalb der *DDX3Y* 5´UTRs konnten, mit dem Programm „WebSat“ (Martins et al., 2009), keine weiteren Mikrosatelliten (Mono- bis Hexanukleotidrepeats), mit einer Komplexität von mindestens drei Repeateinheiten, kartiert werden.
Für *DDX3X* konnte nur in der langen 5´UTR mit Start in T-TSS-III ein Pentanukleotidrepeat der Sequenz „TCCCC“ (BAC-Pos. 29.943 - 29.957; 15 bp), mit einer Komplexität von drei Repeateinheiten innerhalb der distalen Pyrimidin-reichen Sequenz, identifiziert werden.

Als potentielle Faktoren einer Translationskontrolle der *DDX3X* Transkripte in männlichen Keimzellen konnten zusammenfassend verschiedene Sequenz- und Strukturelemente identifiziert werden: die 5´UTR Länge, zwei uORFs, das 5´UTR-Splicing, die Pyrimidin-reiche Region, die 5´TOP-Sequenz und die Adenin-reiche Sequenz. Mit dem (G,C)-Gehalt und der Kartierung mehrerer potentieller G-Quadruplexstrukturen (QGRS) konnten zwei Faktoren identifiziert werden, die direkten Einfluß auf die Sekundärstruktur der mRNA nehmen können. In Abb. 4-74 sind diese Faktoren für die Human *DDX3X* 5´UTR graphisch dargestellt. Der überwiegende Teil der hier beschriebenen, potentiell *cis*-regulativen Elemente befindet sich in dem ausschließlich hodenspezifisch transkribierten *DDX3X* 5´UTR-Bereich. Für mehrere Sequenzabschnitte, wie uORF1/YY1-TFBS, PARS, Pyrimidin-reiche Region und die G-Quadruplexsequenzen, sind zudem auch Kontrollfunktionen auf Transkriptionsebene anzunehmen.

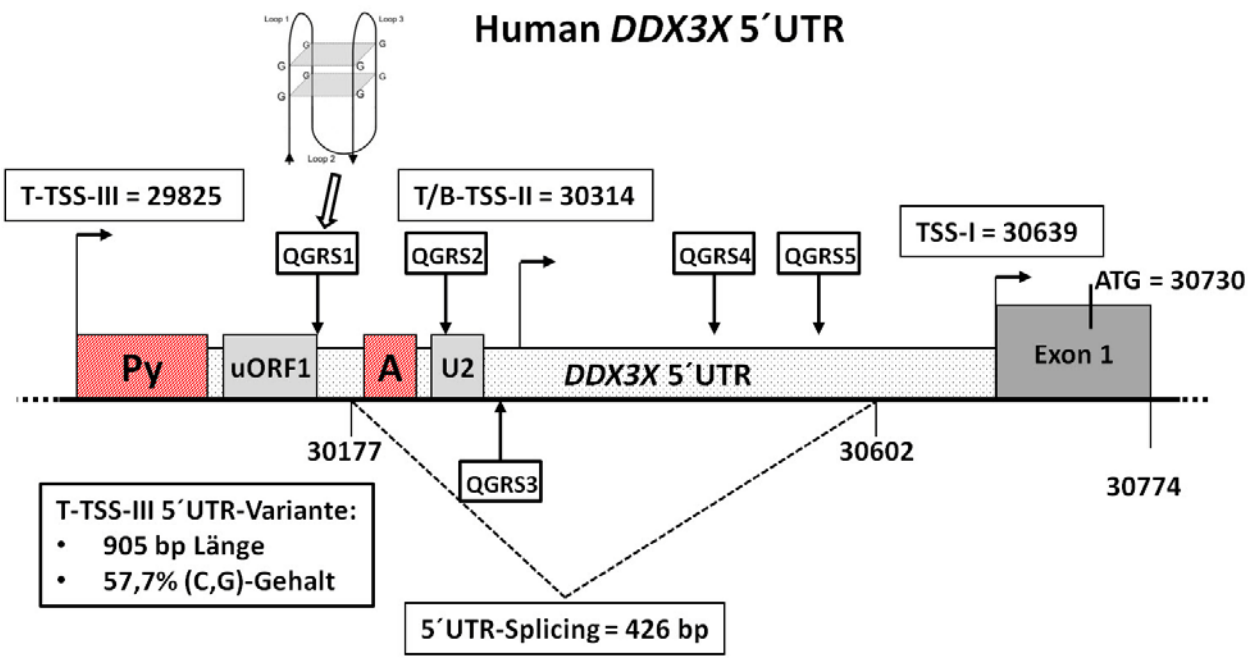


Abb. 4-74: **Schematische Darstellung potentiell *cis*-regulativer Elemente in der *DDX3X* 5´UTR.**
Die Positionsangaben beziehen sich auf den X-chromosomalen BAC-Klon „RP13-13A3". Py = Pyrimidin-reiche Region (BAC-Pos. 29.826 - 29.962; 137 bp), A = PARS „Polypurin (A)-reiche Sequenz" (BAC-Pos. 30.186 - 30.226; 41 bp), uORF1 (BAC-Pos. 30.004 - 30.081; 78 bp, 25AS), U2 = uORF2 (BAC-Pos. 30.257 - 30.292; 36 bp, 11AS), QGRS = „Quadruplex forming G-rich Sequences". G-Quadruplexstrukturmodell verändert nach Wong & Huppert (2009). Für weitere Erklärungen siehe Text.

4.11.3. *In-silico* Vorhersage der optimalen *DDX3X* und *DDX3Y* RNA-Sekundärstruktur nach Zuker-Algorithmus

In diesem Abschnitt sollen mittels Zuker-Algorithmus und dem Programm „mfold" die optimalen Sekundärstrukturen (2D) für die *DDX3Y* und *DDX3X* 5´UTR-RNA-Sequenzen berechnet werden. Der mfold-Algorithmus analysiert zur Strukturvorhersage die minimale freie Energie einer RNA zu verschiedenen thermodynamischen Faltungsmodellen (Mathews et al., 1999; Zuker, 2003). Das Strukturmodell mit der kleinsten freien Energie ΔG sollte die stabilste und folglich optimale Sekundärstruktur für eine gegebene RNA-Sequenz aufweisen (Batey et al., 1999). Der Zuker-Algorithmus berechnet die Sekundärstruktur einer RNA nach zwei Grundsätzen. Erstens soll die Anzahl der Basenpaarungen maximiert und zweitens dabei gleichzeitig die freie Energie minimiert werden, wobei helikale RNA-Duplex-Regionen negative (stabilisierende) und Schleifen positive (destabilisierende) Energien aufweisen. Helikale Strukturen geben also Energie ab bei der Bindung der Basenpaare und Schleifen nehmen Energie auf. Die minimale freie Energie ΔG einer RNA-Struktur wird in kcal/mol angegeben. Zu unterscheiden ist hierbei zwischen der freien Energie für ein gesamtes RNA-Molekül und der freien Energie einzelner Strukturen in einem RNA-Molekül, die zur Gesamtenergie addiert werden. Diese RNA-Strukturmodelle können ebenfalls zur Vorhersage der Translations-

fähigkeit einer RNA verwendet werden. In der Nähe der 5´-Cap-Struktur einer mRNA wird der Scanning-Mechanismus der kleinen ribosomalen Untereinheit bereits von Hairpin-Strukturen mit freien Energien von -30 kcal/mol deutlich beeinträchtigt, da das Binden der Translations-maschinerie erschwert wird. Weiter stromabwärts gelegene Strukturen in der 5´UTR mit freien Energien von -50 kcal/mol und niedriger, können die Ribosomenfunktion gänzlich unterbinden, da das Entwinden der RNA durch die Helikase eIF4A nicht mehr korrekt möglich ist (Kozak, 1989; Pickering & Willis, 2005).

In Tab. 4-44 werden die freien Energien ΔG jeweils für die gesamte *DDX3Y* und *DDX3X* 5´UTR-Sequenzen, sowie jeweils auch für die Einzelstrukturen mit der negativsten freien Energie ΔG , aufgeführt.

Tab. 4-44: Vergleich der freien Energien ΔG, der 5´UTR-Länge und des (C,G)-Gehaltes der verschiedenen *DDX3Y* und *DDX3X* 5´UTR-Varianten.

Positionsbeschreibungen beziehen sich für *DDX3Y* auf BAC 475I1 und für *DDX3X* auf BAC RP13-13A3. In den Spalten 4 und 5 sind die Einzelstrukturen mit den größten freien Energien, mit Position innerhalb der 5´UTR, angegeben. Die ΔG-Werte wurden von mfold für 37ºC und 1M NaCl (keine divalenten Ionen) berechnet.

Gen	TSS & 5´UTR-Variante	Gesamt ΔG 5´UTR (kcal/mol)	ΔG negativste Einzelstruktur (kcal/mol)	Position der Struktur mit max. neg. ΔG	Länge 5´UTR (nts)	(C,G)-Gehalt (in %)
DDX3Y	(54.051) T-TSS-II Ta-1b	-171,04	-38,9	C ^{52}/G ^{134} – U ^{58}/G ^{119}	575	49,04
	(54.051) T-TSS-II Ta-1a	-190,47	-38,5	C ^{52}/G ^{134} – U ^{58}/G ^{119}	652	49,08
	(54.051) T-TSS-II Tb-1b	-203,84	-38,5	C ^{51}/G ^{135} – U ^{58}/G ^{119}	674	48,07
	(54.317) T-TSS-I Ta-1b	-92,79	-18,5	G ^{24}/C ^{226} – A ^{34}/U ^{217}	309	52,4
	(54.317) T-TSS-I Ta-1a	-112,22	-18,5	G ^{24}/C ^{226} – A ^{34}/U ^{217}	386	51,8
	(54.317) T-TSS-I Tb-1b	-123,73	-18,5	G ^{24}/C ^{226} – A ^{34}/U ^{217}	408	50
	(54.911) TSS-I-(T)	-70,75	-20.8	U ^{53}-A ^{207} – G ^{163}-C ^{197}	235	52,8
	(55.032) TSS-Ia	-24,85	-12.1	A ^{46}-U ^{83} – U ^{50}-A ^{79}	114	49,1
	(55.104) TSS-Id	-5,2	-5.4	U ^{6}-A ^{39} – A ^{9}-U ^{36}	42	50
DDX3X	(29.825) T-TSS-III	-320,46	-22,3	U ^{73}/A ^{423} – C ^{79}/G ^{420}	905	57,7
	(29.825) T-TSS-III Splice	-148,48	-17.7	C ^{133}-G ^{258} – G ^{142}-C ^{248}	479	61,0
	(30.314) T/B-TSS-II	-165,16	-19,3	A ^{226}/U ^{355} – G ^{229}/C ^{353}	416	59,9
	(30.639) TSS-I	-30,13	-12.6	G ^{22}-U ^{29}	91	59,3

Die hodenspezifische *DDX3Y* 5´UTR Transkriptvariante Ta-1b mit Start in T-TSS-II weist ein Gesamt-ΔG von -171,04 kcal/mol auf. 52 Nukleotide stromabwärts der 5´Cap-Struktur ist eine Einzelstruktur mit einer freien Energie von -38,9 kcal/mol zu finden (Tab. 4-44; Abb. 4-75). Dieser Wert entspricht den aus der Literatur (ΔG ≤ -30 kcal/mol) beschriebenen Strukturen in 5´Cap Nähe mit Translationsrepressionspotential. In Abb. 4-75 ist zu sehen, daß mfold eine

Vielzahl an Stems und Loops für die *DDX3Y* 5´UTR vorhergesagt. Es fällt auf, daß sich die meisten *DDX3Y* uORF ATGs, uATGs und Stopcodons in Schleifen (Loops) und Ausbuchtungen (Bulges) befinden. Der Vergleich der *DDX3Y* T-TSS-I und T-TSS-II 5´UTRs zeigt, daß jeweils die Ta-1b Splicevariante ein weniger negatives Gesamt-ΔG aufweist (Tab. 4-44). Verglichen mit den Splicevarianten Ta-1a und Tb-1b sollte das Haupt-Spliceprodukt Ta-1b daher insgesamt weniger stabile Strukturen ausbilden und, entsprechend diesen Strukturmodellen, effizienter translatierbar sein. Die *DDX3Y* 5´UTR-Varianten mit Start in „TSS-Ia bis d" sollten entsprechend der mfold-Analyse keiner 5´UTR Translationskontrolle durch Sekundärstrukturen unterliegen. Dagegen ist für die *DDX3Y* 5´UTR-Variante mit Start in TSS-I-(T) eine Translationskontrolle über die mRNA-Struktur, auf Grund der freien Energie ΔG von -70,75 kcal/mol, möglich.

Für die ungesplicte *DDX3X* T-TSS-III 5´UTR wurde eine Struktur mit einer freien Energie ΔG von -320,46 kcal/mol vorhergesagt (Tab. 4-44, Abb. 4-76). Die stabilste Einzel-Struktur innerhalb der ungesplicten 5´UTR liegt in einer Duplex-Region mit einem internen Bulge und hat eine freie Energie von -22,30 kcal/mol (Abb. 4-76). Dieser Bereich liegt innerhalb einer größeren Duplex-Struktur, in der Sequenzen aus dem distalen Pyrimidin-reichen Bereich (BAC-Pos. 29.857 – 29.905) mit Purin-reichen Sequenzabschnitten (BAC-Pos. 30.244 – 30.294) interagieren. Die Pyrimidin-reichen Sequenzen, die nicht Teil der Duplex-Struktur sind, bilden eine große instabile Loop-Struktur mit geringer freier Energie aus (Abb. 4-76). Eine isolierte Analyse der Pyrimidin-reichen Region weist keine Selbstkomplementarität auf. Es bildet sich eine große Loop-Struktur mit -3 kcal/mol.

Die *DDX3X* A-reiche Region bildet ebenfalls eine Loop-Region aus (Abb. 4-76). Eine Duplex-Struktur entsteht für die A-reiche Sequenz nicht, da kein längerer T-Stretch bzw. U-Stretch innerhalb dieser *DDX3X* 5´UTR-Variante vorkommt. Die ATGs der zwei *DDX3X* uORFs sind jeweils in einem Bulge vorhergesagt, was zuvor auch für die *DDX3Y* uORFs beobachtet werden konnte. Die Lage der *DDX3X* und *DDX3Y* uORF ATGs in offenen mRNA-Strukturen könnte einen Hinweis darauf geben, daß diese von *trans*-Faktoren erkannt und gebunden werden können. Die vorhergesagte Gesamtstruktur der ungesplicten *DDX3X* 5´UTR-Sequenz zeigt also, daß die 5´UTR ein hohes Potential zur Bildung von Sekundärstrukturen besitzt, die Einfluß auf die Translationseffizienz haben könnten (Abb. 4-76).

Die hodenspezifisch gesplicte *DDX3X* 5´UTR-Transkriptvariante mit Start in T-TSS-III weist eine maximale freie Energie ΔG von -148,48 kcal/mol auf (Tab. 4-44). Für diese *DDX3X* 5´UTR-Variante werden ebenfalls lange Stem-Sequenzen vorhergesagt. Diese bilden sich zwischen der 5´ Py-Region und einer ~ 100 bp langen Sequenz, die direkt vor dem 5´-Splicingsignal liegt. Eine Loop-Struktur innerhalb der Py-Sequenz, wie in der ungesplicten *DDX3X* 5´UTR, ist nicht

zu identifizieren. Die stabilste Einzel-Struktur mit einer freien Energie von -17,7 kcal/mol ist innerhalb der Stem-Struktur, die durch die Py-Region gebildet wird, lokalisiert.

Zwischen der ungesplicten und gesplicten *DDX3X* T-TSS-III 5´UTR-Variante scheint ein Zusammenhang zwischen 5´UTR-Länge und freier Energie zu bestehen, da längere RNA-Sequenzen mehr Sekundärstrukturen ausbilden können (Tab. 4-44). Die fast doppelt so lange ungesplicte *DDX3X* 5´UTR weist auch eine mehr als doppelt so niedrige freie Energie∆G auf.
Im Vergleich zur *DDX3X* T/B-TSS-II Variante ist aber auch zu sehen, daß nicht nur die 5´UTR-Länge die freie Energie beeinflußt, sondern auch der jeweilige *DDX3X* 5´UTR-Sequenz-abschnitt einen variablen Einfluß auf die Struktur nehmen kann. Die *DDX3X* T/B-TSS-II 5´UTR ist kürzer und hat einen geringeren (G,C)-Gehalt als die gesplicte *DDX3X* T-TSS-III 5´UTR, weist aber dennoch eine negativere freie Energie ∆G auf.
Für die *DDX3X* TSS-I 5´UTR sollte mit einer freien Energie von -30,13 kcal/mol keine Translationsrepression auf Grund der mRNA-Struktur auftreten.

Die Unterschiede in 5´UTR-Länge und (C,G)-Gehalt der *DDX3Y* und *DDX3X* 5´UTR-Sequenzen scheinen deutlich Einfluß auf die freie Energie ∆G und somit die mRNA-Struktur zu nehmen (Tab. 4-44).
Ein hoher (C,G)-Gehalt in der 5´UTR führt aber nicht generell zu einer niedrigen freien Energie ∆G. So wurde für die 84 Nukleotid lange beta-Actin 5´UTR mit mfold eine kaum negative freie Energie ∆G von -12 kcal/mol, trotz des hohen (C,G)-Gehalts von 76,2%, berechnet. Von beta-Actin ist bekannt, daß es in vielen Geweben exprimiert wird und die 5´UTR keiner Translationskontrolle unterliegt (Dmitriev et al., 2007).

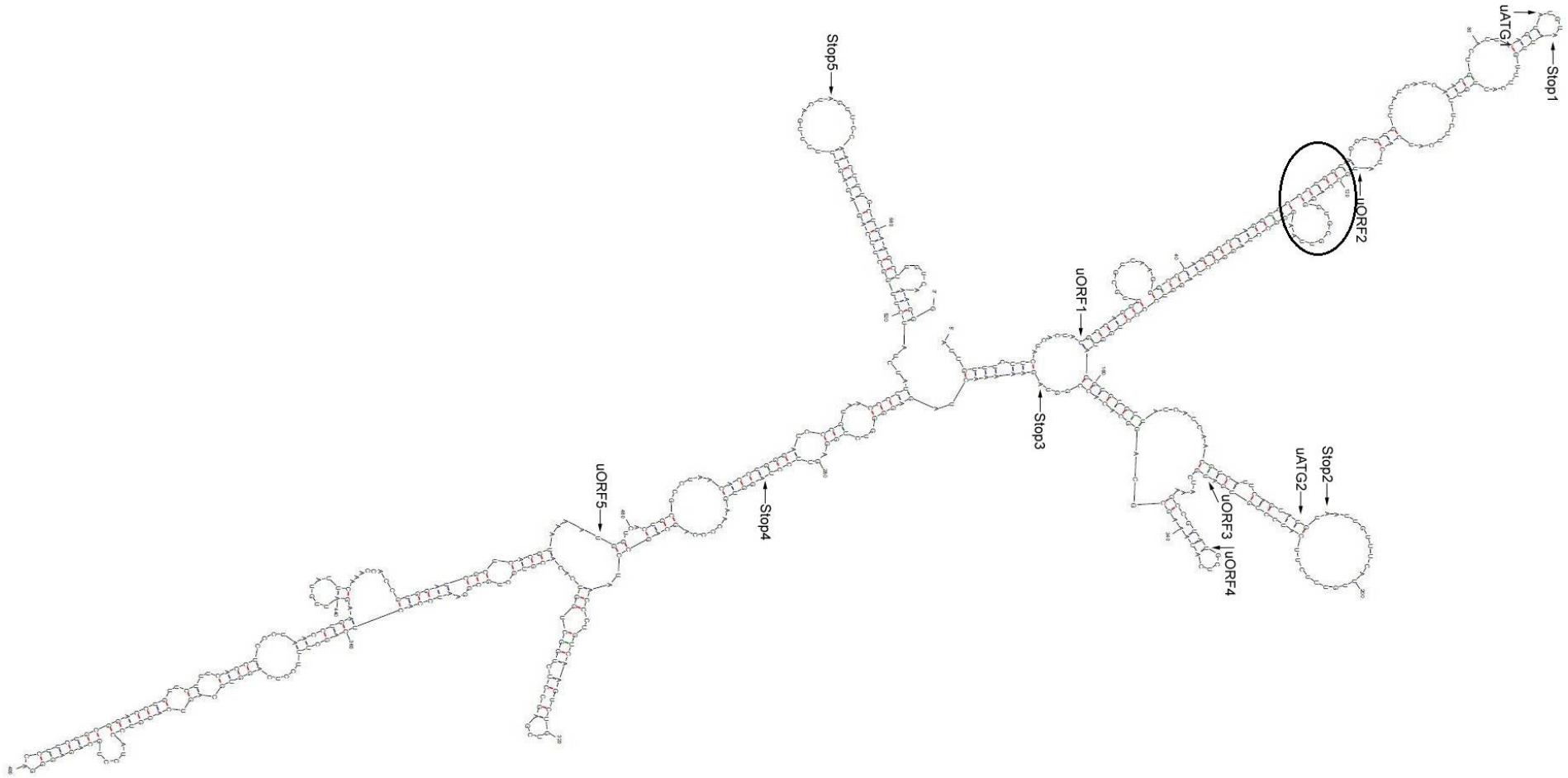


Abb. 4-75: Sekundärstrukturvorhersage für *DDX3Y* 5´UTR (Ta-1b Variante) mit Start in BAC-Position 54.051 im *MSY2-2* Repeat (T-TSS-II).

Für die gesamte 575 Nukleotide lange 5'UTR Sequenz wurde mit mfold (Vers. 3.2) eine freie Energie von ΔG -171,04 kcal/mol vorhergesagt. uORFs und uATGs sind durch entsprechende Beschriftungen markiert. Einzelstruktur größter freier Energie (-38,9 kcal/mol) ist mit einem Kreis markiert. Die ΔG -Werte wurden von mfold für 37ºC und 1M NaCl (keine divalenten Ionen) berechnet.

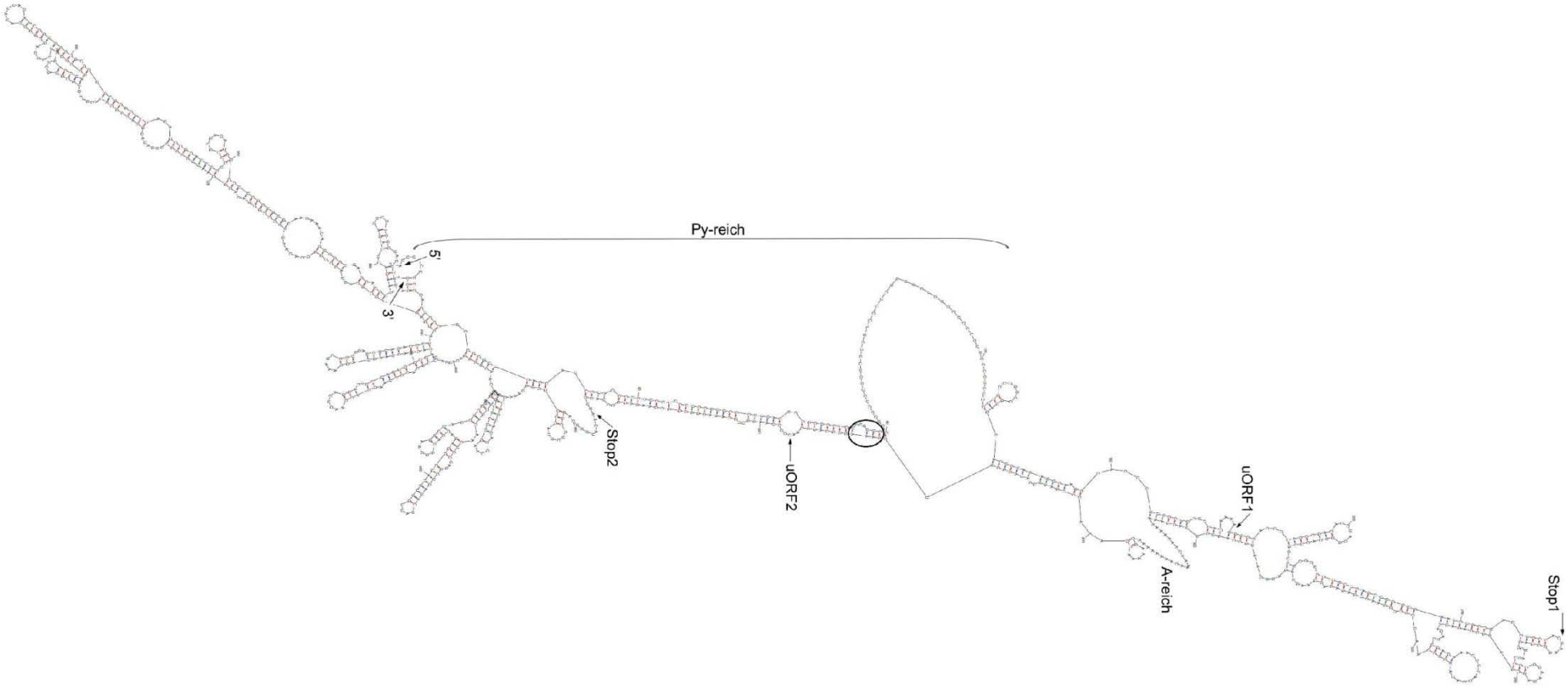


Abb. 4-76: Sekundärstrukturvorhersage für *DDX3X* 5´UTR mit Start in BAC-Position 29.825 (T-TSS-III).

Für die gesamte 905 Nukleotide lange 5´UTR Sequenz wurde mit mfold (Vers. 3.2) eine freie Energie ΔG von -320,46 kcal/mol vorhergesagt. uORFs, PARS und Py-Region sind durch entsprechende Beschriftungen markiert. Einzelstruktur größter freier Energie (-22,3 kcal/mol) ist mit einem Kreis markiert. Die ΔG -Werte wurden von mfold für 37ºC und 1M NaCl (keine divalenten Ionen) berechnet.

Nach Evaluation der *in-silico* Daten wurden die verschiedenen *DDX3X* und *DDX3Y* 5´UTR-Varianten in das von Davuluri et al. (2000) vorgestellte CART-Modell („Classification and regression tree“) zur Klassifizierung humaner 5´UTR-Sequenzen eingeordnet (Abb. 4-77). Anhand dieses CART-Modells sollen Vorhersagen zur Translationsfähigkeit einer 5´UTR-Sequenz getroffen werden können. Aus einer Gruppe von zwölf Faktoren wählten Davuluri et al. (2000) vier Variablen (5´TOP-Sequenz, freie Energie∆G der 5´UTR-Sekundärstruktur, 5´UTR Länge, uATGs) aus, die den größten Einfluß auf die Translationsfähigkeit – in Bezug auf die 5´UTR – einer mRNA aufwiesen. Durch die Bewertung der vier Kriterien können drei unterschiedliche Transkriptklassen (I - III) gebildet werden, die in unterschiedlicher Weise reguliert werden und generell ein unterschiedliches Translationspotential besitzen.

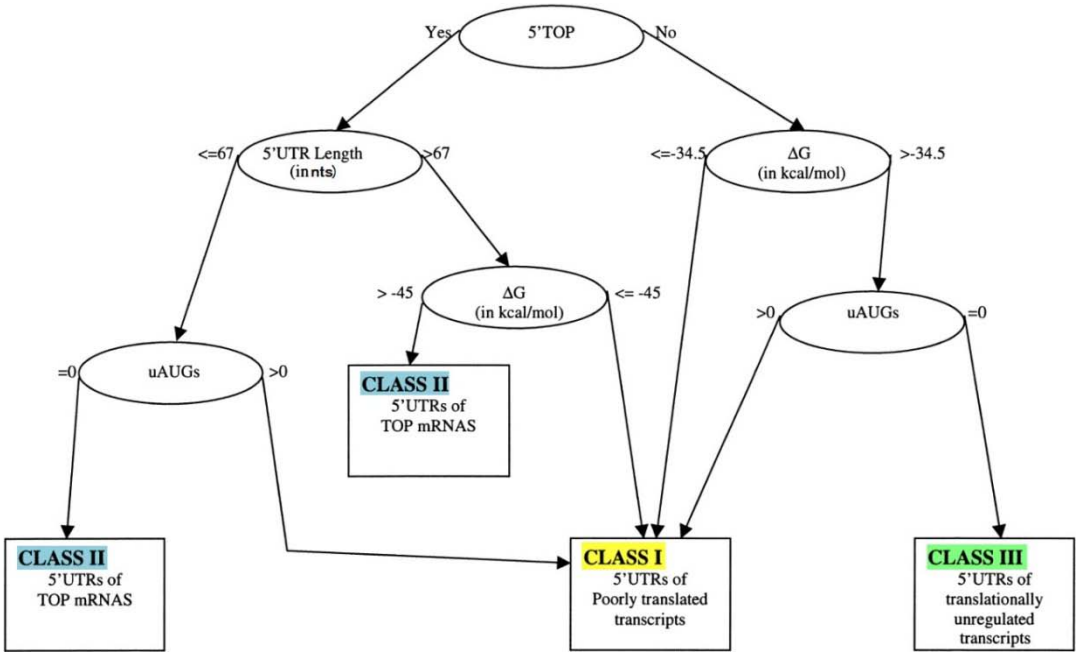


Abb. 4-77: CART-Modell zur Einordnung humaner 5´UTR Sequenzen.
Von zwölf getesteten Variablen in dem CART-System, haben die vier in der Graphik gezeigten Faktoren (5´TOP Anwesenheit, Sekundärstruktur, 5´UTR Länge, uATGs) potentiell den größten Einfluß auf die Translation einer mRNA (nach Davuluri et al., 2000).

Die „Klasse I“ besteht aus Transkripten, die auf Translationsebene durch lange 5´UTRs, uATGs/uORFs und Sekundärstrukturen reguliert werden und dadurch häufig eine geringe Translationseffizienz aufweisen. Die „Klasse II“ besteht aus Transkripten mit 5´TOP-Struktur und kurzen 5´UTRs, zu denen z.B. mRNAs ribosomaler Proteine zählen. Diese Transkripte werden u.a. in Abhängigkeit von Nährstoffangebot und Zellwachstum auf Translationsebene reguliert. In die „Klasse III“ werden Transkripte mit kurzen, unstrukturierten 5´UTRs eingeordnet, die keine uATGs oder uORFs aufweisen. Eine 5´UTR-Länge von ca. 80 Nukleotiden wurde von Kozak (1991) als optimale Länge für eine effiziente Translationsinitiation beschrieben. Eine sehr kurze Leadersequenz wirkt dagegen reprimierend (Kozak, 1991), was mit dem ~ 30 Nukleotide langen Bereich zusammenhängt, der von einem Ribosom belegt wird (Mata et al., 2005). Die resultierenden Proteinmengen werden bei „Klasse III“-Transkripten nicht über die

Translationskontrolle, sondern bereits über die Transkriptionskontrolle und somit über die Menge der vorhandenen Transkripte reguliert (Davuluri et al., 2000). Die ubiquitär exprimierten *DDX3Y* und *DDX3X* TSS-I Transkripte, mit 5´UTR-Längen zwischen 114 und 42 Nukleotiden und freien Energien∆G zwischen -30,13 kcal/mol und -5,2 kcal/mol, würden demnach in die Klasse III einzugliedern sein. Alle anderen in Tab. 4-44 aufgeführten *DDX3Y* und *DDX3X* 5´UTR-Varianten sind der Klasse I zuzuordnen.

4.11.4. Vertikaler Vergleich potentiell *cis*-regulativer Elemente der *DDX3X* und *DDX3Y* 5´UTR-Sequenzen

Das ClustalW-Sequenzalignment der Human *DDX3X* 5´UTR mit den 5´UTR-Sequenzen von acht Säugetierspezies ist im Anhang einzusehen (Abb. A-4). Folgende genomische Sequenzen wurden für das Alignment verwendet: für Schimpanse der BAC-Klon „CH251-3A20“[54], für Gorilla X-chromosomale Sequenzen[55], für Orang-Utan X-chromosomale Sequenzen[56], für *Macaca* X-chromosomale Sequenzen[57], für *Callithrix* BAC-Klon „CH259-42H12“[58], für den Mausmaki X-chromosomale Sequenzen[59], für die Maus BAC-Klon „RP23-158L19“[60] und für das Rind der genomischen X-Contig „004.81“[61].

uORF1:

In Tab. 4-45 wird der Human *DDX3X* uORF1 in den sechs untersuchten Primatenspezies (Schimpanse, Gorilla, Orang-Utan, *Macaca*, *Callithrix* und Mausmaki), sowie Rind und Maus auf Konserviertheit der uORF-Länge, des Abstandes zum Haupt-ATG, des ATG-Kontextes und der Sequenzhomologie verglichen. Für die Spezies mit verfügbaren Y-chromosomalen Sequenzen (Schimpanse, *Macaca*, *Callithrix*, Rind, Maus) wurde daneben der vertikale Vergleich des homologen *DDX3Y* uORF5 in Tab. 4-45 eingefügt. In acht von neun Spezies (außer Mausmaki) ist ein *DDX3X* uORF1 ATG an homologer Position in den genomischen *DDX3X/Ddx3x* Sequenzen identifizierbar (Tab. 4-45). Bei den Primaten ist nicht der gesamte *DDX3X* uORF1 homolog (außer bei Mensch, Schimpanse u. Gorilla), sondern variiert in Länge und Sequenzkonserviertheit. Orang-Utan und Rhesusaffe haben deutlich längere *DDX3X* uORF1 Sequenzen, was reprimierende Auswirkungen auf die Translationsfähigkeit haben könnte. Beim Mausmaki fällt das uORF1 ATG durch Transversion im konservierten ATG weg.

54 Ptro: GenBank Acc.No. AC183643.2; Pos. 465 - 1.640.
55 Ggor: Ensembl Assembly gorGor3, Datenbankvers. 61.3d; Pos. 42.084.176 - 42.085.350.
56 Ppyg: Ensembl Assembly PPYG2, Datenbankvers. 61.1g; Pos. 41.919.443 - 41.920.638.
57 Mmul: Ensembl Assembly MMUL 1.0, Datenbankvers. 61.10p; Pos. 39.101.003 - 39.102.191.
58 Cjac: GenBank Acc.No. AC193908.1; Pos. 97.841 - 96.717.
59 Mmur: Ensembl Assembly micMur1, Datenbankvers. 61.1f; scaffold_4539; Pos. 63.071 - 64.204.
60 Mmus: GenBank Acc.No. AL833805.11; Pos. 89.571 - 90.583.
61 Btau: GenBank Acc.No. NW_001501904.1; Pos. 83.588 - 82.419.

Tab. 4-45: Vertikaler und lateraler (horizontaler) Vergleich des Human *DDX3X* uORF1 und *DDX3Y* uORF5 in genomischen Säugersequenzen.

Der Abstand von *DDX3X* uORF1 zum Haupt-ATG bezieht sich in der Human-Sequenz auf die ungesplicte *DDX3X* 5´UTR. Zwei Punktmutationen in *DDX3X Microcebus murinus* (Mmur) verkürzen den uORF1. Das uORF1 ATG ist bei Mmur nicht positionshomolog zum Human uORF1 ATG. Für *B. taurus* und *M. musculus* konnten in Kapitel 4.5.3. keine 5´verlängerten *Ddx3x* Transkripte ermittelt werden, die ein homologes uORF1 beinhalten würden. Die *DDX3X* uORF1 Homologie bezieht sich immer auf die 78 bp lange genomische Sequenz. Ein *DDX3X* 5´UTR Alignment aller untersuchten Säuger befindet sich im Anhang (Abb. A-4).

Für *DDX3Y* wird die Splicevariante Exon-Ta-1b zur Bestimmung des Abstands verwendet, da nicht-Human Primaten nur diese Variante exprimieren. Für *Callithrix jacchus DDX3Y* Transkripte mit Exon-Ta-1b 5´UTR-Splicing konnte ein *Callithrix*-spezifisches uATG in Exon-Ta festgestellt werden. Dieses beginnt 12 Nukleotide nach dem uORF5 Stopcodon. Für dieses in-frame uATG ist kein Stopcodon vor dem *DDX3Y* Haupt-ATG vorhanden, so daß ein verlängerter DDX3Y ORF entsteht. Der *Callithrix DDX3Y* Leserahmen wird um 18 Nukleotide (6 AS = MKAVKG) am Protein N-Term verlängert. In den dargestellten uORF1/uORF5 ATG-Kontextsequenzen ist das ATG rot hervorgehoben, die kritischen Positionen -3 und +4 sind grün markiert. Unterschiede zur Human-Sequenz sind unterstrichen. Die Sequenzhomologie bezieht sich auf die DNA-, bzw. RNA-Sequenz. Abstand = erstes Nukleotid nach uORF bis incl. A von ATG. k.A. = keine Angabe möglich, da keine genomische Y-Sequenz verfügbar.

Spezies		***DDX3X* uORF1**				***DDX3Y* uORF5**				Sequenz-Homologie zw. Human X/Y uORF
		Länge uORF1	Abstand zu Haupt-ATG	Sequenz-Kontext uORF ATG (gccRccATGG)	Sequenz-Homologie zu Human uORF1	Länge uORF5	Abstand zu Haupt-ATG in Ta-1b Variante	Sequenz-Kontext uORF ATG (gccRccATGG)	Sequenz-Homologie zu Human uORF5	
Primaten	*Homo sapiens*	78 nts (25 AS)	649 nts	CACAAAATGG	100%	72 nts (23 AS)	30 nts	CGTAAAATGG	100%	48,7%
	Pan troglodytes	78 nts (25 AS)	620 nts	CACAAAATGG	100%	72 nts (23 AS)	30 nts	CATAAAATGG	98%	47,4%
	Gorilla gorilla	78 nts (25 AS)	619 nts	CACAAAATGG	98,7%	k.A.				
	Pongo pygmaeus	114 nts (37 AS)	615 nts	CACAAAATGG	93%	k.A.				
	Macaca mulatta	156 nts (51 AS)	567 nts	CACAAAATGG	92,3%	72 nts (23 AS)	30 nts	CGTAAAATGG	94%	26,9%
	Callithrix jacchus	66 nts (21 AS)	626 nts	CACAAAATGG	87,2%	66 nts (21 AS)	30 nts	CTTAAAATGG	76,4%	62,1%
	Microcebus murinus	[42 nts (13 AS)]	611 nts	GATCACATGA	75,6%	k.A.				
Artiodactyla	*Bos taurus*	[264 nts (--- AS)]	419 nts	CACAAAATGG	65,4%	66 nts (21 AS)	119 nts	TACGAAATGA	59,7%	13,3%
Rodentia	*Mus musculus*	[291 nts (--- AS)]	329 nts	CACAAGATGG	46,2%	kein Transkript vorhanden mit Exon-T; ATG nicht vorhanden in genomischer Sequenz				

Positionshomolog zur Human uORF1-Sequenz entsteht in der Mausmaki Sequenz durch Transition eines GTG-Codons, innerhalb des Human uORF1, ein neues uATG. Das Mausmaki-spezifische uORF, mit einer Länge von 42 bp (13 AS), wird hier uORF3 genannt. Die dazugehörige Kozak-Sequenz hat einen schwachen Translationskontext (-3 = C; +4 = A). In der Rind und Maus 5´ *Ddx3x* Sequenz konnten Position und uORF1 ATG-Kontext ebenfalls homolog zur Humansequenz identifiziert werden. In Rind und Maus ist dieser uORF allerdings ohne Funktion für die Translation, da dieser nur in der genomischen Sequenz identifiziert wurden und keine 5´ verlängerten Rind oder Maus *Ddx3x* Transkripte mit diesem uORF nachgewiesen werden konnten.
In dem hodenspezifischen Rind *Ddx3x* Transkript mit Start in Position T-TSS-II sind zwei Rind-spezifische uORFs entstanden. Der erste Rind uORF, *Ddx3x* uORF6 genannt, ist 12 Nukleotide stromabwärts der T-TSS-II Position lokalisiert und 48 Nukleotide (15 AS) lang. Die dazugehörige Kozak-Sequenz ist in einem starken Translationskontext (-3 = A; +4 = G). Der zweite Rind-spezifische uORF, *Ddx3x* uORF7 genannt, beginnt 8 Nukleotide stromabwärts des ersten Rind uORFs und ist ebenfalls 48 Nukleotide (15 AS) lang. Die dazugehörige Kozak-Sequenz hat einen adäquaten Translationskontext (-3 = G; +4 = A). Beide Rind *Ddx3x* uORFs überlappen. In dem Maus *Ddx3x* Transkript ist kein uORF oder uATG vorhanden.

Die Länge und die Sequenzhomologie des Human *DDX3Y* uORF5 in Exon-Ta scheint in den untersuchten Spezies weniger variabel, als für den *DDX3X* uORF1, zu sein. In Schimpanse, *Macaca* und *Callithrix* ist dieser uORF zu 100% positionshomolog zu finden. Einzig in der *Callithrix* Sequenz ist eine 6 bp lange Deletion entstanden, die aber den uORF Leserahmen nicht verändert. Innerhalb der Primaten ist für *DDX3Y* uORF5 eine hohe Konserviertheit des ATG-Kontextes feststellbar. Für das Rind konnte innerhalb des Rind Exon-T´s ein 66 bp (21 AS) langer uORF kartiert werden, der teilweise mit Human uORF5 überlappt und sich im selben Leserahmen befindet. Die dazugehörige Kozak-Sequenz hat einen adäquaten Translations-kontext (-3 = G; +4 = A). In Maus konnte kein 5´ verlängertes Transkript mit uORF festgestellt werden. In der genomischen Sequenz ist ebenfalls kein homologes ATG in der Maus-Sequenz festzustellen.

uORF2:

Der Human *DDX3X* uORF2 ist, außer beim Orang-Utan mit derselben 26 bp lange Sequenz-duplikation, bei den nicht-Human Primaten nicht konserviert wiederzufinden (Tab. 4-46).
In der 5´UTR-Sequenz von *Macaca DDX3X* ist durch Mutation zweier Nukleotide ein neues uATG (TTA zu ATG) entstanden, wodurch ein neuer *Macaca*-spezifischer uORF, hier uORF4 genannt, mit 183 Nukleotiden (60 AS) entsteht. Dieser befindet sich 24 Nukleotide strom-abwärts des Stopcodons TGA von *Macaca DDX3X* uORF1. Die Kozak-Sequenz von *Macaca* uORF4 ist stark (-3 = G; +4 = G). Auch bei *M. murinus* entsteht durch Substitution eines GTG

Tripletts ein neues uATG, wodurch ein zweiter Mausmaki-spezifischer uORF, hier uORF5 genannt, mit 126 bp (41AS) Länge entsteht. Das neue uATG befindet sich 26 bp stromaufwärts des Human uORF2 ATG. Die dazugehörige Kozak-Sequenz in der Mausmaki-Sequenz hat einen schwachen Translations-kontext (-3 = T; +4 = C).

Einfluß von uATGs auf den DDX3X und DDX3Y ORF (N-terminal):

Eine Überprüfung aller *DDX3X* und *DDX3Y* 5´UTR-Varianten auf eine potentielle Veränderung der DDX3X und DDX3Y Proteinleserahmen ergab einzig für das *Callithrix jacchus DDX3Y* T-TSS-I Transkript (Abb. 4-15 u. 4-19, Kapitel 4.4.1) eine mögliche N-terminale Verlängerung des DDX3Y Proteins. Es handelt sich um eine Verlängerung von 18 nts, bzw. 6 Aminosäuren, die durch eine mögliche Translationsinitiation an einem in-frame uATG mit schwachem Kontext „GATTGAATGA“ entstehen kann. Das uATG befindet sich zwischen dem *DDX3Y* uORF5 Stop-codon und der Exon-Ta 5´GT-Splicestelle. Eine Analyse mit dem Protein-Domänen Analyse-Programm “Prosite” ergab kein bekanntes Muster für das Hexa-Peptid „MKAVKG“. Welchen Einfluß und welche Funktion diese DDX3Y Proteinvariante in *Callithrix jacchus* hat, bzw. ob diese mögliche Proteinvariante tatsächlich *in-vivo* synthetisiert wird, ist nicht bekannt.

Tab. 4-46: Vertikaler Vergleich von Human *DDX3X* uORF2 in verschiedenen Säugerspezies.
Der Abstand von *DDX3X* uORF2 zum Haupt-ATG bezieht sich in der Human-Sequenz auf die ungesplicte 5´UTR. Bei *Macaca* und Mausmaki ist uORF2 nicht homolog zum Human uORF2. Für *B. taurus* und *M. musculus* konnte jeweils kein *DDX3X* Transkript ermittelt werden, welches vor Human uORF2 beginnt. Für das Rind wurde in Kapitel 4.5.2 ein Transkriptstart vergleichbar zur Human T/B-TSS-II Position vorgestellt. In diesem Rind Transkript sind zwei uORFs kartierbar, die nicht den Human uORFs entsprechen. Rind uORF1 ist unten gezeigt. In der genomischen Maus-Sequenz ist ein nicht-homologer uORF kartierbar. Für Maus konnte keine verlängerte 5´UTR festgestellt werden, so daß dieser Maus uORF ohne Funktion bleibt. Für *Macaca*, Mausmaki, Rind und Maus werden die uORF Angaben in eckige Klammer gesetzt, da diese uORFs nicht homolog zum Human uORF2 sind. In den dargestellten uORF2 ATG-Kontextsequenzen ist das ATG rot hervorgehoben, die kritischen Positionen -3 und +4 sind grün markiert. Unterschiede zur Human-Sequenz sind unterstrichen. Die Sequenzhomologie bezieht sich auf die DNA-, bzw. RNA-Sequenz. Abstand = erstes Nukleotid nach uORF bis incl. A von ATG. Ø = uORF2 nicht vorhanden.

Spezies		***DDX3X* uORF2**			
		Länge uORF2	Abstand zu Haupt-ATG	Seq.-Kontext uORF ATG (gccRccATGG)	Sequenz-Homologie zu Human uORF2
Primaten	*Homo sapiens*	36 nts (11 AS)	438 nts	GAGCGAATGC	100%
	Pan troglodytes	Ø	Ø	Ø	Ø
	Gorilla gorilla	Ø	Ø	Ø	Ø
	Pongo pygmaeus	36 nts (11 AS)	440 nts	GAGCAAATGA	86%
	Macaca mulatta	[183 nts (60 AS)]	[361 nts]	[CCAGTGATGG]	Ø
	Callithrix jacchus	Ø	Ø	Ø	Ø
	Microcebus murinus	[93 nts (30 AS)]	[345 nts]	[AACTGGATGC]	Ø
Artiodactyla	*Bos taurus*	[48 nts (15 AS)]	[361 nts]	[CGGAAAATGG]	Ø
Rodentia	*Mus musculus*	[24 nts (7 AS)]	[262 nts]	[TTCTTTATGA]	Ø

QGRS:

Der Vergleich der potentiellen Human *DDX3X* QGRSs mit den non-Human Primatenspezies zeigt, daß Human *DDX3X* QGRS4 (zwischen uORF2 und 5´-Splicesignal) in 7 von 7 untersuchten Primatenspezies konserviert vorliegt (Tab. 4-47). Die *DDX3X* QGRS1 (direkt nach uORF1), QGRS3 (direkt nach uORF2) und QGRS5 (vor der 5´ Splice-Akzeptorstelle) kommen jeweils in 6 von 7 Primatenspezies konserviert vor (Tab. 4-47). *DDX3X* QGRS2, die im Human uORF2 entsteht, ist humanspezifisch. Für den Orang-Utan konnte keine QGRS in der homologen *DDX3X* uORF2 Sequenz kartiert werden. Für Ptro, Ggor und Mmul wurde mit QGRS-Mapper eine neue *DDX3X* QGRS in der uORF2 Sequenzregion kartiert. Sie entsteht vermutlich durch Deletion der Human 26 Nukleotid-Duplikation, im Bereich von uORF2, neu in diesen Spezies.

Zur Übersicht sind in Tab. 4-48 die vorgestellten potentiellen *cis*-regulativen Elemente der Human *DDX3X* 5´UTR im Vergleich zu den acht oben genannten Säugetierspezies zusammengefaßt. Innerhalb der Primaten konnte insgesamt eine hohe Konserviertheit dieser Elemente festgestellt werden. Die *DDX3X* 5´UTR-Längen, der (C,G)-Gehalt, die Pyrimidin-reiche Region sowie PARS und uORF1 scheinen mit kleineren Abweichungen evolutionär sehr stabil zu sein. Konservierte regulative *cis*-Funktionen für die Transkriptions- und Translations-kontrolle von *DDX3X* sind daher wahrscheinlich. Das keimbahnspezifische *DDX3X* 5´UTR-Splicing ist als einziger untersuchter Faktor humanspezifisch. Für Rind und Maus konnten diese *cis*-regulativen Elemente nicht gefunden werden, da die meisten dieser Elemente im hodenspezifischen 5´UTR-Abschnitt stromaufwärts der *DDX3X* T/B-TSS-II Position lokalisiert sind. Einzig der Human DDX3X QGRS5 konnte in Rind homolog kartiert werden.

Tab. 4-47: Vergleich der potentiellen Human *DDX3X* 5´UTR G-Quadruplex Sequenzen in *DDX3X* 5´UTRs von non-Human Primaten.

Gezeigt sind die fünf potentiellen Human G-Quadruplex Sequenzen wie in Tabelle 3 vorgestellt. Parameter: Max. Länge = 30 Nukleotide, Min. G-Gruppen = 2, Loop-Größe von 0 bis 36 Nukleotide, nicht überlappende QGRS. G-Gruppen sind unterstrichen. Ptro = *Pan troglodytes*, Ggor = *Gorilla gorilla*, Ppyg = *Pongo pygmaeus*, Mmul = *Macaca mulatta*, Cjac = *Callithrix jacchus*, Mmur = *Microcebus murinus*. Die G-Gruppen sind in den jeweiligen Sequenzen unterstrichen. Ø = keine homologe QGRS-Sequenz in non-Human Primaten.

Nr.	Human BAC-Position	QGRS-Länge	QGRS-Sequenz	G-Score	Homologie
1	30.082	27	GGGGAGGCTACCCGCGGAAGAGCGAGG	16	Ptro, Ggor, Ppyg, Mmul, Mmur
2	30.265	15	GGTTGGAAAGGGGGG	20	Ø
3	30.306	25	GGCGTCTTAGCCGGTGGAAAGCGGG	12	Ptro, Ggor, Ppyg, Mmul, Cjac
4	30.452	30	GGACTGCTAGGGGCGACAGGACTAAGTAGG	20	Ptro, Ggor, Ppyg, Mmul, Cjac, Mmur
5	30.543	24	GGAATTGCGGTGTGAGAGGGAGGG	17	Ptro, Ggor, Mmul, Cjac, Mmur

Tab. 4-48: Zusammenfassung und Speziesvergleich potentieller *cis*-Faktoren in der Human *DDX3X* 5´UTR.

Die non-Human Primaten 5´UTR-Längen wurden für Schimpanse, *Macaca* und *Callithrix* in Kapitel 4.4.1 anhand der RT-PCR Ergebnisse identisch zu Human in Position *DDX3X* T-TSS-III festgelegt. Für Gorilla, Orang-Utan und Mausmaki wird auf Grund der bisher ermittelten Primatenergebnisse ebenfalls ein *DDX3X* Transkriptstart in der T-TSS-III Position angenommen. *Ddx3x* 5´UTR-Längen für Rind und Maus wurden in Kapitel 4.5.2. anhand der RT-PCR Ergebnisse ermittelt. Für Rind konnte eine 5´UTR-Verlängerung festgestellt werden. Als Rind *Ddx3x* Transkriptstart wurde die T-TSS-II Position vorgeschlagen, die identisch zu Human *DDX3X* T/B-TSS-II lokalisiert ist. Für Maus konnte eine offensichtlich gewebespezifische *Ddx3x* TSS-I 5´UTR-Verlängerung ermittelt werden. k.A. = keine Angabe möglich, da kein Gewebe verfügbar war. Ø = nicht vorhanden. uORF Angaben in eckigen Klammern bedeuten, daß keine Homologie zu den Human uORFs besteht. Py = Pyrimidin. PARS = Polypurin (A)-reiche Sequenz. TISU = Translation Initiator of Short 5´UTR. QGRS = Quadruplex forming G-rich Sequences.

Spezies	Länge distale Py-reiche Sequenz	Länge PARS	Länge uORF1	Länge uORF2	5´UTR Splicing, Splice-Länge	Pot. TISU-Element (uORF1 ATG)	Anzahl pot. QGRS	(C,G)-Gehalt	Max. Länge der 5´UTR
Hsap	137 bp	41 bp	78 bp	36 bp	Ja, 426 bp	Ja	5	57,7%	905 bp
Ptro	137 bp	38 bp	78 bp	Ø	Ø	Ja	5	57.88%	876 bp
Ggor	138 bp	37 bp	78 bp	Ø	k.A.	Ja	6	57,94%	875 bp
Ppyg	133 bp	41 bp	114 bp	36 bp	k.A.	Ja	4	56,65%	902 bp
Mmul	134 bp	62 bp	156 bp	[uORF4 = 183 bp]	Ø	Ja	5	57,64%	897 bp
Cjac	95 bp	35 bp	66 bp	Ø	Ø	Ja	6	56,71%	827 bp
Mmur	123 bp	22 bp	[uORF3 = 42 bp]	[uORF5 = 126 bp]	k.A.	Ø	6	56,62%	846 bp
Btau	Ø	Ø	[uORF6 = 48 bp	[uORF7 = 48 bp]	Ø	Ø	3	58,1%	420 bp
Mmus	Ø	Ø	Ø	Ø	Ø	Ø	1	60,14%	143 bp

4.11.5. *In-vitro* Translations-Assay zur Messung der Translationseffizienz der Human *DDX3X* 5´UTR-Varianten

Im Folgenden werden die *in-silico* erhobenen Daten und *cis*-Elemente auf ihren potentiellen Einfluß der DDX3X 5´UTR-Varianten untersucht. Durch die Nutzung eines gekoppelten *in-vitro* Transkriptions/Translations-Assays (TNT Quick Coupled Transcription/Translation Systems, Promega, # L1170) soll der translationale Einfluß der *DDX3X* 5´UTRs, mit den drei Startpositionen in T-TSS-III, T/B-TSS-II und TSS-I, in einem Luziferase-Reportergensystem untersucht werden.

Die Transkription, in diesem für Säugetier-RNAs optimierten eukaryotischen *in-vitro* System, ist ausschließlich abhängig von der T7-Promotersequenz (5´-TAATACGACTCACTATAGGG-3´), die von der T7 RNA-Polymerase erkannt wird. D.h., daß die langen Bereiche der *DDX3X* 5´UTRs mit Start in T-TSS-III, die *in-vivo* auch als Promoter funktionieren können, in diesem *in-vitro* System ausschließlich als 5´UTR-Sequenz, und nicht als alternativer, interner Promoter erkannt werden. Die Beschreibung des Vektoraufbaus, aus pUC18-Vektor und Firefly-Luziferase [*luc2*], ist im Methoden-Teil (Kapitel 3) einsehbar.

Um die unterschiedlichen *DDX3X* 5´UTR-Varianten auf ihren Translationseinfluß zu analysieren, wurde eine Serie rekombinanter Plasmide, mit verschiedenen Bereichen der Human *DDX3X* 5´UTR generiert. Diese sind in Abb. 4-78 schematisch dargestellt und in Tabelle 4-49 beschrieben. Es wurden 14 5´UTR Konstrukte für diesen Assay generiert, die in vier Kategorien einzuteilen sind.
Die erste „Kategorie grün“ (H1 - H4) besteht aus den vier 5´UTR-Varianten, die nativ in Human Hodengewebe zu detektieren sind. Die Human-Konstrukte-1 (H1) und -2 (H2) repräsentieren die zwei 5´UTR-Varianten mit Start in der *DDX3X* T-TSS-III Position. H2S enthält die gesplicte 5´UTR-Variante. Konstrukt H3 beinhaltet die *DDX3X* TSS-II 5´UTR. Das 5´Ende dieses Konstruktes H3 wurde auf die BAC-Position 30.280 gelegt, der Startposition des 5´RACE-Klons (Kapitel 4.2), da zum Zeitpunkt dieser Experimentserie die kartierte *DDX3X* T/B-TSS-II in BAC-Position 30.314 noch nicht eindeutig bestätigt war. Wie in den nativen *DDX3X* T/B-TSS-II 5´UTR-Varianten sind in Konstrukt H3 ebenso keine uORFs vorhanden. Das TSS-I Transkript (H4) startet aus technischen Gründen in BAC-Position 30.637, die unmittelbar an der ubiquitären *DDX3X* TSS-I Position (BAC-Pos. 30.639) liegt.
Die zweite „Kategorie gelb“ (H5 - H8) besteht aus artifiziellen 5´UTR-Sequenzvarianten, die zur Unterscheidung eines potentiellen Einflusses der zwei *DDX3X* uORFs generiert wurden. Das Konstrukt H5 beginnt in BAC-Position 30.222, stromabwärts des 5´Splicesignals, beinhaltet den uORF2 und wurde generiert, um einen potentiellen Einfluß von uORF2 isoliert von uORF1

betrachten zu können. Konstrukt H6 startet in BAC-Position 29.959 und beinhaltet beide uORFs, aber nicht die Pyrimidin-reiche Region am distalen Ende der *DDX3X* 5´UTR. Mit diesem Konstrukt soll, im Vergleich zu H1, H7 und H8, evaluiert werden, ob die beiden uORFs alleine, oder nur in Kombination mit der distalen Pyrimidin-reichen Region einen Einfluß auf die Translationsfähigkeit der *DDX3X* T-TSS-III Transkriptvarianten ausüben. Die artifizielle 5´UTR-Variante H7 beinhaltet nur den distalen 5´UTR-Bereich bis einschließlich uORF2. Konstrukt H8 wurde zur isolierten Analyse des uORF1 und des distalen 5´UTR-Bereiches generiert und endet direkt vor dem 5´-Splice-Signal.

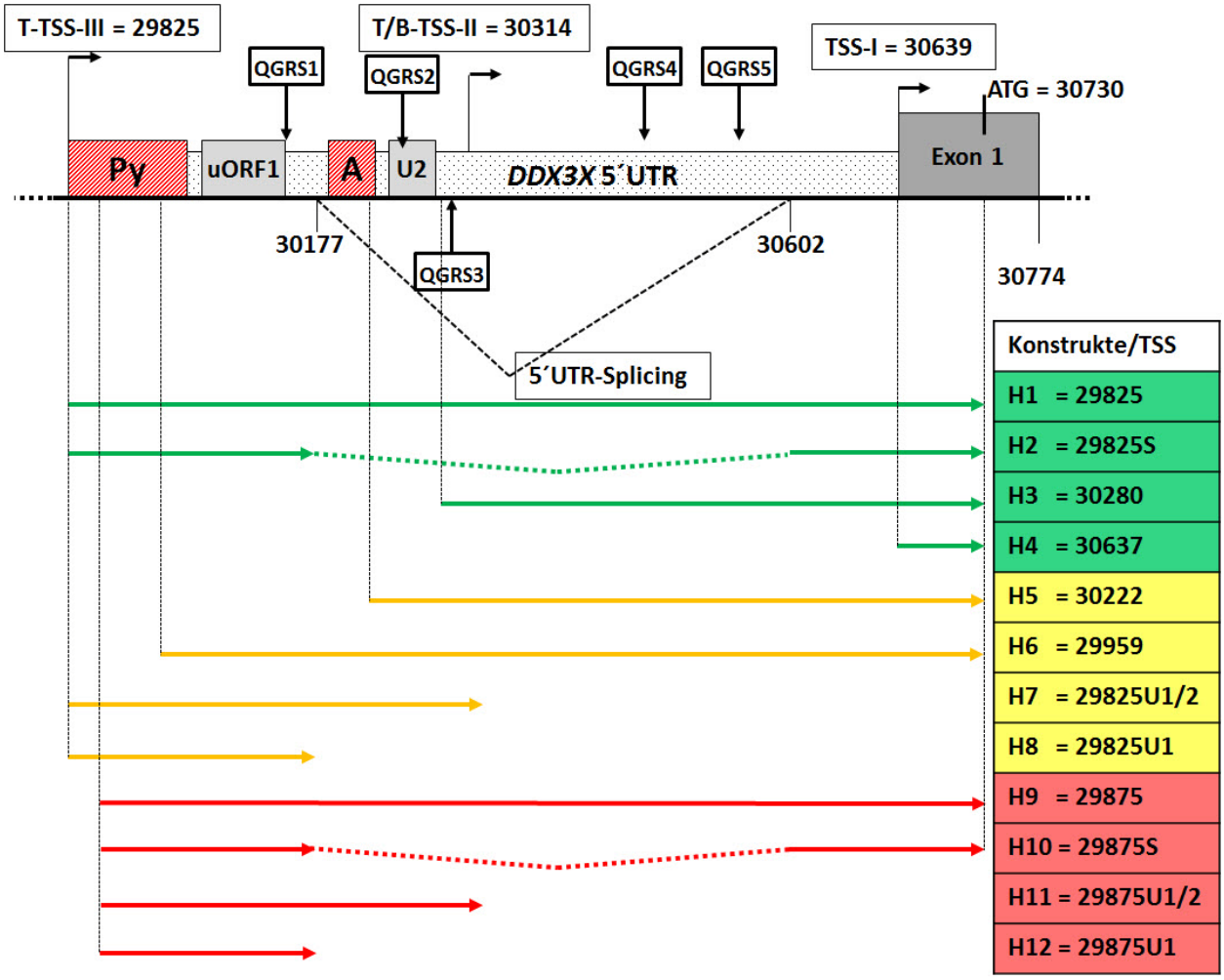


Abb. 4-78: Schematische Darstellung der verschiedenen Human *DDX3X* 5´UTR Konstrukt-varianten.

Die gezeigten 5´UTR Sequenzen wurden jeweils in das pUC18-*Luc2* Plasmid kloniert. Die klonierten 5´UTR Sequenzen sind in dem Schema durch Pfeile repräsentiert und in Bezug zu den *in-silico* kartierten pot. Kontroll-elementen dargestellt. Die Konstrukte sind in vier Kategorien eingeteilt. „Kategorie grün“ repräsentiert die nativ vorkommenden *DDX3X* 5´UTR-Transkriptvarianten. Die „Kategorie gelb“ besteht aus künstlich generierten *DDX3X* 5´UTR-Abschnitten zur Evaluation isolierter 5´UTR Elemente. „Kategorie rot“ besteht aus vier Konstrukten, die in der Pyrimidin-reichen Region, aber nicht in der kartierten Position T-TSS-III, mit einer potentiellen 5´TOP-Sequenz beginnen. Die vierte „Kategorie blau“ mit den „uORF1-Mutationskonstrukten“ ist hier nicht gezeigt (siehe Tab. 4-49). Die genaue Beschreibung der klonierten Sequenzen erfolgt in untenstehender Tabelle und im Text. Abkürzungen in schematischer *DDX3X* 5´UTR Darstellung: Py = Pyrimidin-reiche Region, A = Polypurin (A)-reiche Sequenzen (PARS), U2 = uORF2, QGRS = „Quadruplex forming G-rich Sequences“. Abkürzungen in rechtsstehender Tabelle: S = Splicing; U1/2 = Konstrukt beinhaltet uORF1 und uORF2. U1 = Konstrukt beinhaltet nur uORF1, nicht uORF2.

Tab. 4-49: Beschreibung der verschiedenen *DDX3X* 5´UTR Konstrukte für den *in-vitro* Translations-Assay.

Die 5´UTR Sequenzen sind entsprechend der schematischen Abb. 4-78 und der Beschreibung im Text in vier, farblich unterschiedliche, Kategorien unterteilt. Als Merkmale der verschiedenen Konstrukte wurden die TSS-Position, die 5´UTR Länge, die uORFs, die 5´TOP Sequenz, der (C,G)-Gehalt und die freie Energie∆G angegeben. U1/2 = Konstrukt enthält uORF1 und -2. U1 = Konstrukt beinhaltet nur uORF1, Ø = nicht vorhanden.

Spezies	Konstrukt-bezeichnung	TSS	Position im BAC RP13-13A3	Länge der klonierten 5´UTR	uORFs	5´TOP	(C,G)-Gehalt 5´UTR in %	∆G nach mfold in kcal/mol
Homo sapiens	H1	T-TSS-III	29.825-30.729	905 nts	Nr.1 + 2	ø	57,63	-320
	H2 (Splicing)	T-TSS-III	29.825-30.176 + 30.603-30.729	479 nts	Nr. 1	ø	61,0	-148
	H3	TSS-II	30.280-30.729	450 nts	ø	ø	58,9	-172
	H4	TSS-I	30.637-30.729	93 nts	ø	ø	60,2	-27
	H5	---	30.222-30.729	508 nts	Nr. 2	ø	58,3	-196
	H6	---	29.959-30.729	771 nts	Nr.1 + 2	ø	56,42	-270
	H7 (U1/2)	T-TSS-III	29.825-30.331	507 nts	Nr.1 + 2	ø	56,2	-156
	H8 (U1)	T-TSS-III	29.825-30.176	352 nts	Nr. 1	ø	60,8	-87
	H9	---	29.875 - 30.729	855 nts	Nr.1 + 2	ja	57,1	-309
	H10 (Splicing)	---	29.875-30.176 + 30.603-30.729	429 nts	Nr.1	ja	60,1	-136
	H11 (U1/2)	---	29.875-30.331	457 nts	Nr.1 + 2	ja	54,9	-146
	H12 (U1)	---	29.875-30.176	302 nts	Nr. 1	ja	59,6	-87
	H13 = H1 Mut. uORF1	T-TSS-III	29.825-30.729	905 nts	Nr.2	ø	57,63	-325
	H14 = H6 Mut. uORF1	---	29.959-30.729	771 nts	ø	ø	56,42	-270

Die 5´UTR-Sequenzen der Konstrukte H7 und H8 sind so gewählt, daß mind. 50 Nukleotide zwischen dem uORF (H7 = uORF2; H8 = uORF1) und dem Luziferase-Translationsstartcodon vorhanden sind und somit eine effiziente Reinitiation am *luc2* ATG ermöglicht wird.

Die Konstrukte der dritten „Kategorie rot“ (H9 - H12) beginnen an der BAC-Position 29875, die in den NCBI-Datenbanken als historisch erstes potentielles 5´Ende des *DDX3X* Transkriptes NM_001356 zu finden war. Mit dieser Konstrukt-Kategorie kann der potentielle Einfluß einer 5´TOP-Sequenz am Transkriptbeginn ungezielt untersucht werden. Die Konstrukte H9 und H10

sind analog den Konstrukten H1 und H2, nur daß der Sequenzbeginn in BAC-Position 29875 liegt. H11 und H12 sind analog zu den Konstrukten H7 und H8.

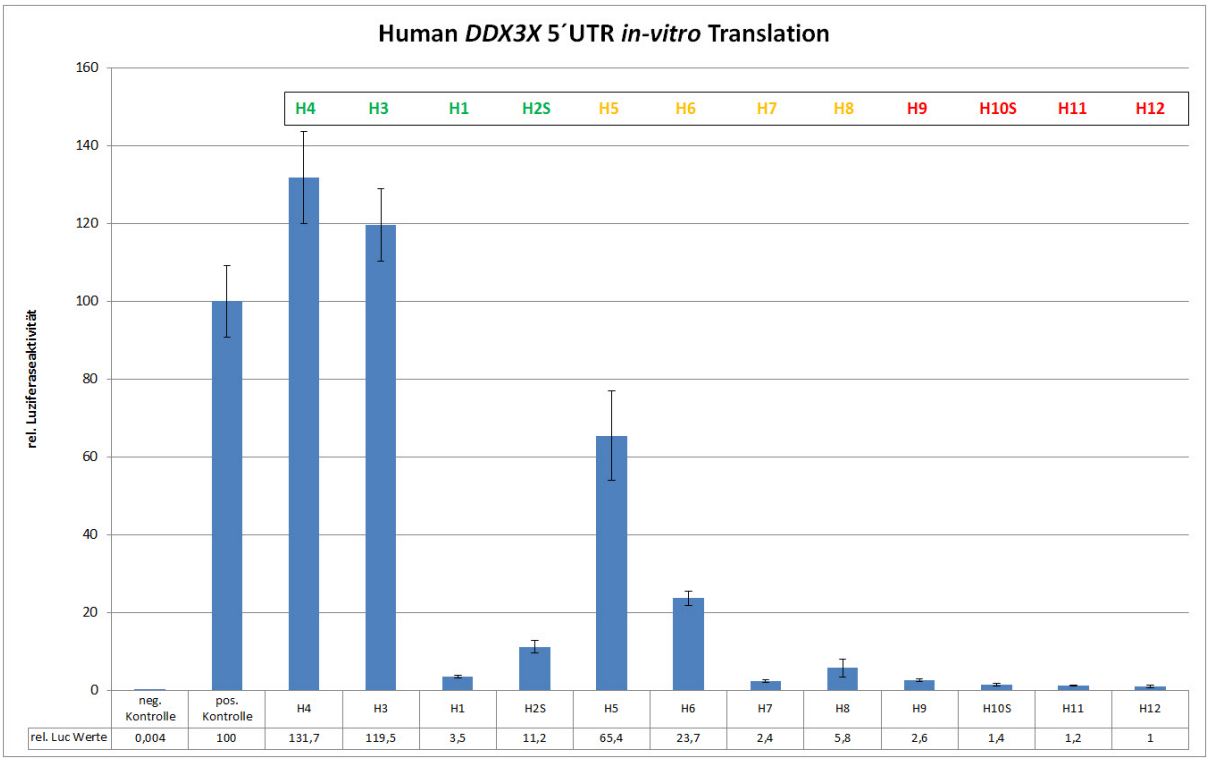
Die vierte „Kategorie blau“ (H13, H14) besteht aus zwei Konstrukten, in denen das ATG des uORF1 per „site-directed Mutagenesis“ zu TAG mutiert wurde. Für diese uORF1-Mutation wurden die Konstrukte H1 und H6 ausgewählt. Die eingeführte Mutation hat keinen Einfluß auf den (C,G)-Gehalt und die freie Energie∆G der klonierten *DDX3X* 5´UTR-Sequenzen. Die Integrität der mutierten 5´UTR Konstrukte und korrekte Integration der Mutation wurden durch Sequenzierung bestätigt. Die Überprüfung der freien Energie∆G ergab für H13 („H1 Mut uORF1“) eine geringfügige Erniedrigung auf -325 kcal/mol. Für H14 („H6 Mut uORF1“) ist keine Veränderung der freien Energie ∆G festzustellen.

Versuchsdurchführung:

Die Ergebnisse des Human *DDX3X* 5´UTR *in-vitro* Translations-Assays sind in Abb. 4-79 und Tab. 4-50 dargestellt. Für die kurze, ubiquitär exprimierte 5´UTR (TSS-I) in Konstrukt H4 konnte eine rel. Luziferaseaktivität von 132% festgestellt werden. Dies bedeutet eine 37,7fache Steigerung gegenüber der *DDX3X* T-TSS-III 5´UTR (H1) in diesem Versuchsaufbau.
Für das Konstrukt H3 (TSS-II) konnte ebenfalls keine Translationsrepression in diesem Versuchsaufbau festgestellt werden. Die rel. Luziferaseaktivität wurde mit 119,5% bestimmt, was im Vergleich zu H1 einer 34fachen Aktivitätssteigerung entspricht. Das Translations-repressionselement, oder -Elemente, sollte dementsprechend in den distalen 455 Nukleotiden der langen *DDX3X* 5´UTR, stromaufwärts von BAC-Position 30.280, lokalisiert sein. Für H3 konnte ebenfalls gezeigt werden, daß die 5´UTR-Länge (450 nts) und der (C,G)-Gehalt (59%) keine negativen Eigenschaften darstellen.
Für die Konstrukte H3 und H4 konnte eine effizientere *in-vitro* Translation, als für das im Promega-Kit gelieferte Kontrollplasmid „Luziferase T7 Kontroll-DNA“, hier „T7-Kontrolle“ genannt, festgestellt werden. Die 5´UTR der T7-Kontrolle ist 47 bp lang und hat einen (C,G)-Gehalt von 59,57%. Zudem stammt das Luziferasegen [*luc*+], des T7-Kontrollvektors, aus der pGL2-Vektorgeneration und ist in der „Codon Usage“ noch nicht für eine Säugerexpression optimiert (Paguio et al., 2005). Die *luc*+ Kozak-Sequenz ist in einem adäquaten Translationskontext (-3 = C; +4 = G). In der 5´UTR der Kontrollsequenz ist zudem ein uORF (57 bp; 18 AS) lokalisiert, der mit dem *luc*+ ORF überlappt. Die dazugehörige Kozak-Sequenz hat allerdings einen schwachen Translationskontext (-3 = C; +4 = C).

Für die lange, durchgehende *DDX3X* 5´UTR (H1) konnte, im Vergleich zu der T7-Kontrolle, nur 3,5% relative Luziferaseaktivität gemessen werden. Die gesplicte *DDX3X* 5´UTR H2 erreichte einen Wert von 11,2% Luziferaseaktivität. D.h., durch das Heraussplicen des uORF2, der

Adenin-reichen Region und der QGRS2 bis 5 ist, im Vergleich zu H1, eine 3,2fache Steigerung der Translationseffizienz möglich.



	neg. Kontrolle	pos. Kontrolle	H4	H3	H1	H2S	H5	H6	H7	H8	H9	H10S	H11	H12
rel. Luc Werte	0,004	100	131,7	119,5	3,5	11,2	65,4	23,7	2,4	5,8	2,6	1,4	1,2	1

Abb. 4-79: Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays.
Die Positivkontrolle „T7-Kontrolle“ ist als Referenz auf 100% gesetzt. Alle experimentellen Luziferasewerte werden in Bezug zur „T7-Kontrolle“ gesetzt. Gezeigt sind die zwölf Konstrukte H1 bis H12 der Kategorien grün, gelb und rot. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-50 aufgeführt.

Mit dem Konstrukt H5 konnte eine rel. Luziferaseaktivität von 65,4% erzielt werden. Dies entspricht einer 1,8fachen Reduktion gegenüber Konstrukt H3. Die 58 nts längere H5 Sequenz, in der uORF2 und QGRS2 lokalisiert sind, führt also zu einer deutlichen Reduktion der *DDX3X* 5´UTR Translationseffizienz. Da 5´UTR-Länge, (C,G)-Gehalt und ΔG der beiden Sequenzen H3 und H5 keine deutlichen Unterschiede aufweisen, kommen somit der uORF2 und die QGRS2 als potentiell reprimierende Elemente in Betracht.
Für das Konstrukt H6, welches am 3´Ende der Pyrimidin-reichen Region beginnt, ist eine rel. Luziferaseaktivität von 23,7% festzustellen. Verglichen mit Konstrukt H5 bedeutet es eine weitere 2,8fache Reduktion der Translationseffizienz. D.h. also, daß die zusätzliche Anwesenheit von uORF1, der A-reichen Region (PARS) und von QGRS1 in H6, im Vergleich zu H5, zu einer weiteren deutlichen Reduzierung der Translationseffizienz führen.

Die geringen Translationsaktivitäten der artifiziellen Konstrukte H7 (2,4%) und H8 (5,8%) weisen ebenfalls darauf hin, daß potentiell Translations-repressive Elemente vor allem im 5´Bereich der *DDX3X* 5´UTR lokalisiert sind. Im Vergleich beider Konstrukte ist zu sehen, daß

eine fast vollständige Translationsrepression durch die gesamte *DDX3X* 5´UTR Sequenz stromaufwärts der T/B-TSS-II Position erreicht wird.

Tab. 4-50: Absolute und relative Werte der *in-vitro* Messung.
Die Positivkontrolle „T7-Kontrolle“ ist als Referenz auf 100% gesetzt. Der Mittelwert der absoluten Werte ergibt sich aus der mindestens 5-fachen Messung eines Translationsansatzes.

Konstrukt	Absoluter *Luc*-Messwert	Relativer *Luc*-Wert in %	Standardabweichung in %
Neg. Kontrolle	167,8	**0,004**	0,0003
Pos. Kontrolle	3.811.256,3	**100**	9,2
H1	134.260,2	**3,5**	0,4
H2S	428.327,3	**11,2**	1,7
H3	4.555.133,5	**119,5**	9,3
H4	5.020.519,7	**131,7**	11,8
H5	2.493.347,6	**65,4**	11,5
H6	901.877,5	**23,7**	1,8
H7 (U1/2)	90.526,3	**2,4**	0,3
H8 (U1)	221.397,9	**5,8**	2,3
H9	99.053,6	**2,6**	0,4
H10S	52.732,1	**1,4**	0,4
H11 (U1/2)	46.866,8	**1,2**	0,03
H12 (U1)	37.948,3	**1**	0,4
H13 (H1 Mut uORF1)	94.693,2	**2,5**	0,3
H14 (H6 Mut uORF1)	393.563,4	**10,3**	1,3

Die Konstrukte H9 bis H12 der „Kategorie rot“, mit 5´UTR Startstelle in BAC-Position 29.875, weisen allesamt eine sehr geringe Translationseffizienz auf, die nochmals geringer ausfallen, als für die entsprechenden Konstrukte H1, H2S, H7 und H8 mit der nativen Startstelle *DDX3X* T-TSS-III in BAC-Position 29.825. Der Vergleich der absoluten Luziferase-Meßwerte in Tab. 4-50 weist darauf hin, daß die 5´TOP Sequenz zu Beginn der 5´UTR einen zusätzlich reprimierenden Effekt auf diese Konstrukte ausübt.

Um zu prüfen, ob uORFs die Translationseffizienz der 5´UTR tatsächlich negativ beeinflussen, wurde in den Konstrukten H1 und H6 das uORF1 ATG zu TAG mutiert. Überraschenderweise führte diese ATG-Mutation zu einer weiteren Reduktion der Translationsfähigkeit dieser Konstrukte (Abb. 4-80). Für H13 („H1 Mut uORF1“) fällt dieser Effekt nur sehr geringfügig aus, da bereits die Translationseffizienz des Konstruktes H1 sehr gering war. Dagegen zeigt das H14 („H6 Mut uORF1“) Konstrukt eine 2,3fache Translationsreduktion im Vergleich zu H6, was in den absoluten Werten in Tab. 4-50 deutlich zu sehen ist.

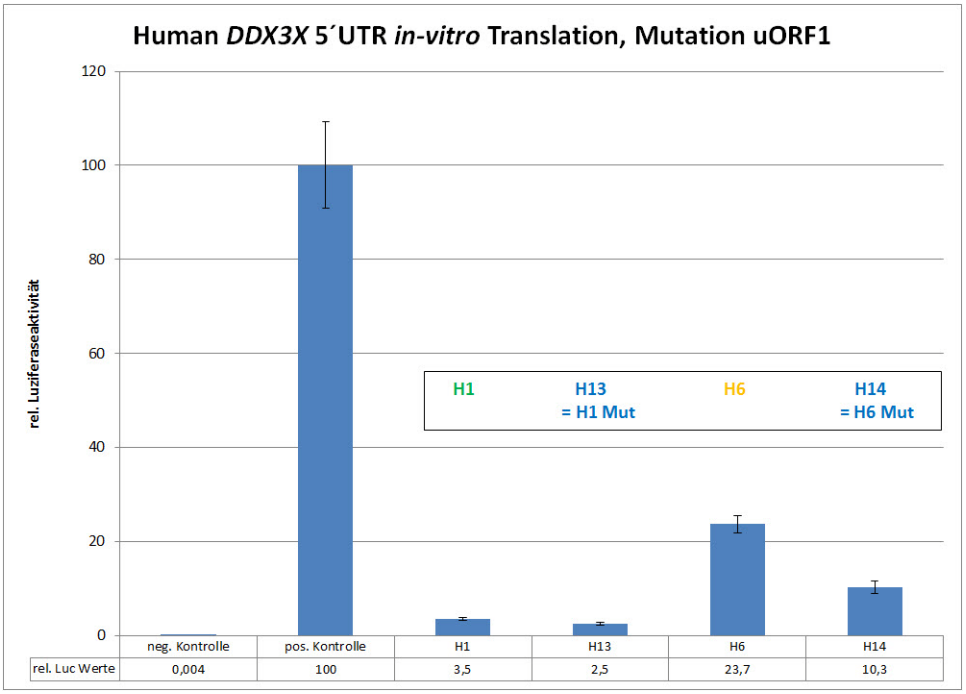


Abb. 4-80: Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays mit Mutation in uORF1.
Die Positivkontrolle „T7-Kontrolle“ ist als Referenz auf 100% gesetzt. Gezeigt sind die zwei Konstrukte H13 („H1 Mut uORF1“) und H14 („H6 Mut uORF1“) der Kategorie blau im Vergleich zu H1 und H6 ohne Mutation. Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-50 aufgeführt.

Zusammenfassend lässt sich also sagen, daß repressive Translations-Kontrollelemente im hodenspezifischen 5´Bereich der *DDX3X* 5´UTR zu suchen sind. Die Funktion des uORF1 konnte in diesem *in-vitro* Translationsansatz nicht abschließend geklärt werden, da mit den zwei angewandten Herangehensweisen, Verkürzung der 5´UTR-Sequenzen und uORF1 Mutationsanalyse, gegenteilige Ergebnisse erzielt wurden. Eine Abhängigkeit vom Kontext der klonierten 5´UTR-Sequenz und der gewählten Methode ist daher anzunehmen. Es sollte nicht außer Acht gelassen werden, daß es sich hierbei um einen *in-vitro* Translationsassay handelt, in dem nur die Einflüsse *cis*-regulativer *DDX3X* 5´UTR-Elemente auf die Grundkomponenten des Translationsapparates untersucht werden. Mögliche positive und negative *trans*-Effekte von *in-vivo* vorhandenen RNA-bindenden Proteinen werden mit diesem Ansatz bewußt nicht erfaßt.

4.11.6. *In-vitro* Translations-Assay zur Messung der Translationseffizienz der *DDX3X* 5´UTR Transkriptvarianten in Primaten

Zur Evaluation, ob die unterschiedliche Translationsfähigkeit verlängerter *DDX3X* 5´UTRs auch in anderen Primatenspezies bereits existiert, wurden die hodenspezifischen *DDX3X* 5´UTR-Verlängerungen von Schimpanse, *Macaca* und *Callithrix* mit dem gleichen funktionellen *in-vitro* Translationsassay studiert. Es wurde jeweils ein *DDX3X* T-TSS-III und ein TSS-II 5´UTR Konstrukt generiert (Tab. 4-51). In Kapitel 4.5.1 fiel bereits auf, daß die Schimpansen und

Makaken *DDX3X* RNA-Expression im distalen Pyrimidin-reichen 5´UTR-Bereich geringer, als für Human, ausfiel (Abb. 4-36, 4-37). Die Schimpansen und Makaken *DDX3X* T-TSS-III 5´UTR Konstrukte wurden daher analog zum Human-Konstrukt H6, mit Start in BAC-Position 29.959 am 3´Ende der Pyrimidin-reichen Region, generiert. Für *Callithrix jacchus* konnte in Kapitel 4.5.1 (Abb. 4-38) eine zu Human vergleichbare Initiation in der Pyrimidin-reichen Region festgestellt werden. Die P2402 Primerposition wurde als Start des *Callithrix DDX3X* 5´UTR-Konstruktes gewählt.

Primaten *DDX3X* 5´UTR-Konstrukte mit Start in der *DDX3X* TSS-I Position wurden auf Grund der hohen Sequenzhomologie im proximalen 5´UTR-Bereich, zwischen den Primatenspezies, nicht erstellt.

Zur Analyse einer potentiellen uORF1 Funktion für die *DDX3X* 5´UTR Translationsfähigkeit, wurde von dem Schimpansen TSS-III Konstrukt P6 ebenfalls eine Variante mit uORF1 ATG Mutation erstellt. Da in der Schimpansen *DDX3X* 5´UTR-Sequenz uORF2 nicht vorhanden ist, wird mit diesem 5´UTR-Konstrukt somit die Translationsfähigkeit, nach Mutation, ohne Anwesenheit eines funktionellen uORFs untersucht.

Tab. 4-51: Beschreibung der verschiedenen non-Human Primaten *DDX3X* 5´UTR Konstrukte für den *in-vitro* Translations-Assay.

Startpositionen der Primaten-Konstrukte beziehen sich auf die entsprechenden Positionen im Human BAC RP13-13A3. Als Merkmale der verschiedenen Konstrukte wurden die TSS-Position, die 5´UTR Länge, die uORFs, die 5´TOP Sequenz, der (C,G)-Gehalt und die freie Energie ∆G angegeben. Ø = nicht vorhanden. P = *Pan troglodytes*, M = *Macaca mulatta*, C = *Callithrix jacchus*.

Spezies	Konstrukt-bezeichnung	TSS	Position im Hsap BAC RP13-13A3	Länge der klonierten 5´UTR	uORF	5´TOP	(C,G)-Gehalt 5´UTR in %	∆G (kcal/mol)
Pan troglodytes	P3	TSS-II	Vgl. 30.284	446 nts	ø	ø	59,42	-168
	P6	---	Vgl. 29.959	742 nts	Nr. 1	ø	56,6	-262
	P6 (Mutation)	---	Vgl. 29.959	742 nts	ø	ø	56.6	-262
Macaca mulatta	M3	TSS-II	Vgl. 30.284	448 nts	ø	ø	58,93	-164
	M6	---	Vgl. 29.959	767 nts	Nr. 1, [Nr. 2]	ø	56,2	-263
Callithrix jacchus	C1	T-TSS-III	Vgl. 29.844	807 nts	Nr. 1	ø	56,5	-268
	C3	TSS-II	Vgl. 30.284	448 nts	ø	ø	58,04	-158

Für das *Callithrix* TSS-III Konstrukt C1 wurde ein rel. Luziferasewert von 3,8% gemessen (Abb. 4-81, Tab. 4-52). Dies ist vergleichbar zu dem Wert, der für das Human-Konstrukt H1 festgestellt wurde. Die Py-Sequenz der langen *Callithrix DDX3X* 5´UTR scheint in diesem Versuchsaufbau ebenfalls einen repressiven Einfluß auf die Translationsfähigkeit der 5´UTR zu bewirken.

Für das Schimpansen-Konstrukt P6 wurde ein Wert von 23,7% rel. Luziferaseaktivität festgestellt (Abb. 4-81). Dies entspricht exakt dem Wert, der für das Human-Konstrukt H6 bestimmt wurde. Da in der Schimpansen *DDX3X* 5´UTR kein uORF2 vorkommt, weist dieses Ergebnis auch darauf hin, daß der Human uORF2 offensichtlich keinen, oder nur einen geringen, Effekt auf die *DDX3X* 5´UTR Translationseffizienz ausübt.

Für das *Macaca*-Konstrukt M6 wurde eine rel. Luziferaseaktivität von 10,4% bestimmt. Im Vergleich zu H6 und P6 entspricht dies einer 2,3fach geringeren Luziferaseaktivität. Ursache könnten hier die *Macaca*-spezifischen, langen uORF1 und uORF4 sein. Zudem haben bei *Macaca* die zwei uORFs (uORF1, uORF4) einen geringen Abstand, was theoretisch eine Reinitiation an uORF4, trotz starken ATG-Kontextes, erschweren oder unterbinden sollte. Da die Beschaffenheit der *DDX3X* 5´UTR uORFs den einzig sichtbaren Unterschied zwischen Human- und Makaken-Sequenz darstellt, erscheint es möglich, daß uORF1 und -4 in *Macaca* eine veränderte Funktion auf die Translationsfähigkeit der 5´UTR-Sequenz ausüben.

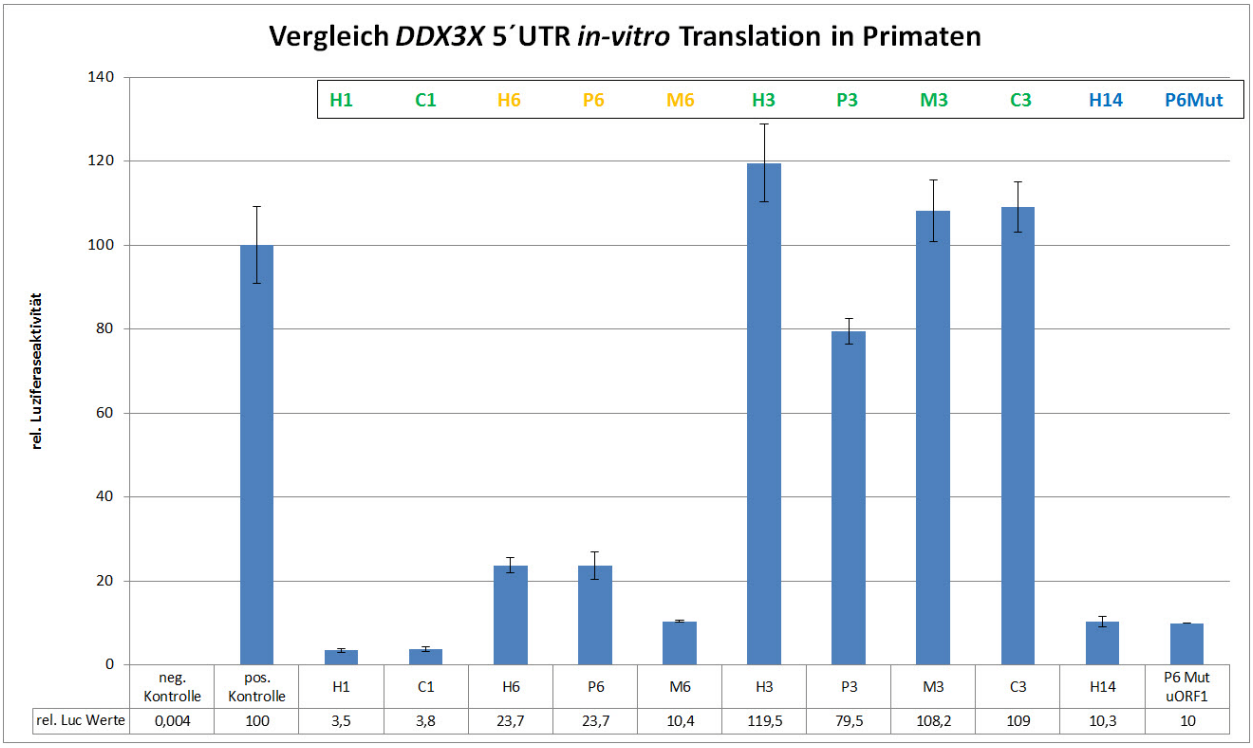


Abb. 4-81: Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays in non-Human Primaten.
Die Positivkontrolle „T7-Kontrolle“ ist als Referenz auf 100% gesetzt. Alle experimentellen Luziferasewerte werden in Bezug zur Positivkontrolle gesetzt. Gezeigt sind die *DDX3X* 5´UTR Konstrukte im Vergleich zwischen Mensch (H), Schimpanse (P), *Macaca* (M) und *Callithrix* (C). Die relativen und gemittelten Luziferasewerte sowie die prozentualen Standardabweichungen sind in Tab. 4-52 aufgeführt.

Für das Schimpansen 5´UTR-Konstrukt P3 wurde eine rel. Luziferaseaktivität von 79.5% festgestellt. Trotz hoher Sequenzhomologie und keiner auffälligen Unterschiede im (C,G)-

Gehalt oder der freien Energie ΔG zu dem Human 5´UTR -Konstrukt H3, fällt die P3 Luziferase-aktivität 1,5fach geringer aus.
Die ermittelten rel. Luziferaseaktivitäten für die *Macaca* M3- und *Callithrix* C3- Konstrukte zeigen keinen deutlichen Unterschied zur Human H3 5´UTR-Sequenz. Die M3- und C3-Luziferasewerte fallen ~ 10% geringer als für H3 aus.

Die eingeführte Mutation in die Schimpansen *DDX3X* 5´UTR im Konstrukt „P6 Mut uORF1" bewirkte eine 2,4fache Reduktion der Luziferaseaktivität im Vergleich zu Konstrukt P6 (Abb. 4-81). Diese negative Veränderung entspricht sehr exakt dem Wert, der für das Human H14 Konstrukt („H6 Mut uORF1") nach Mutation festgestellt wurde. Da dieser repressive Effekt für drei unabhängige Konstrukte (H13, H14, P6 Mut) festgestellt werden konnte, ist nicht von einem Artefakt der Mutations-Methode auszugehen, sondern von einem Effekt, der durch den Austausch von zwei Nukleotiden (ATG zu TAG) entsteht.

Tab. 4-52: Absolute und relative Werte des *DDX3X* 5´UTR *in-vitro* Translationsansatzes mit non-Human Primaten-Konstrukten.
Die Positivkontrolle „T7-Kontrolle" ist als Referenz auf 100% gesetzt. Der Mittelwert der absoluten Werte ergibt sich aus der mindestens 5-fachen Messung eines Translationsansatzes.

Konstrukt	Absoluter *Luc*-Messwert	Relativer *Luc*-Wert in %	Standardabweichung in %
Neg. Kontrolle	167,8	**0,004**	0,0003
Pos. Kontrolle	3.811.256,3	**100**	9,2
P3	3.030.687,0	**79,5**	3
P6	903.766,5	**23,7**	3,2
P6 Mut uORF1	380.076,6	**10**	0,1
M3	4.125.369,8	**108,2**	7,3
M6	396.548,2	**10,4**	0,2
C1	143.691,2	**3,8**	0,5
C3	4.154.787,2	**109**	5,9

Die Ergebnisse der Human und Primaten 5´UTR Translationsanalyse von *DDX3X* 5´UTR-Varianten zeigen, daß die Translationsfähigkeit der verschiedenen *DDX3X* 5´UTR-Varianten innerhalb der vier untersuchten Primatenspezies überwiegend vergleichbar ist. Im Rahmen dieses *in-vitro* Translationsassays konnte dazu eine potentiell Funktion für das konservierte *DDX3X* 5´UTR uORF1 gezeigt werden. Eine entscheidende Funktion für uORF2 konnte insgesamt nicht gezeigt werden. Der deutliche Unterschied zwischen Konstrukt H3 und H5 scheint eher ein Kontext-Effekt der klonierten Sequenzen zu sein, da für das Konstrukt P6 ohne uORF2 keinerlei Veränderung gegenüber dem Human-Konstrukt H6 festgestellt werden konnte. Für die distale Pyrimidin-reiche Sequenz in der *DDX3X* 5´UTR mit Start in T-TSS-III konnte eindeutig eine reprimierende Funktion auf die Translationsfähigkeit gezeigt werden.

5. Diskussion

Die Sequenz des ersten Humangenoms wurde vor zehn Jahren entschlüsselt (Lander et al., 2001; McPherson et al., 2001; Venter et al., 2001; Collins et al., 2003; Frazier et al., 2003; Abdellah et al., 2004; Schmutz et al., 2004; Kidd et al., 2008). Sie ist, wie die Proteinsequenzen der ~ 21.000 Protein kodierenden Gene, in den öffentlichen Datenbanken von NCBI und ENSEMBL abrufbar (Flicek et al., 2011; Sayers et al., 2011).
Die Einfachheit mit der die genetische Information in der DNA-Sequenz durch nur vier Nukleotidbasen gespeichert wird fasziniert, da nicht nur die Proteininformationen im Triplettcode abgelegt werden, sondern durch die Abfolge der vier Nukleotidbasen auch ein strukturelles Muster in der DNA vorgegeben wird. Dieses Muster kann, besonders in genregulativen Sequenzen die Chromatinstruktur mitgestalten, was Auswirkungen auf die Expression der Gene hat. Wesentliche Einsichten in die Evolution der Genregulation können durch die vergleichende Genomik erhalten werden.

Es ist bekannt, daß besonders Gene, die für Fertilität und Reproduktion verantwortlich sind, erhöhten Evolutionsraten und somit einem hohen Selektionsdruck zur Erhaltung der Art unterliegen (Wu & Davis, 1993; Torgerson et al., 2005; Qvarnström & Bailey, 2009). Viele dieser Fertilitätsgene sind auf dem männlichen Y-Chromosom lokalisiert. Auf Grund der fehlenden Rekombination der Geschlechtschromosomen seit ~ 130 Mio. Jahren, konnten ursprünglich homologe Aminosäuresequenzen und regulative Sequenzen evolvieren und Mann-spezifische Eigenschaften annehmen (Wu & Davis, 1993; Goetting-Minesky & Makova, 2006; Ellegren & Parsch, 2007; Hedrick, 2007; Williams & Carroll, 2009).
Das in dieser Arbeit untersuchte gonosomale Genpaar *DDX3X*/*DDX3Y* weist eine Homologie der Aminosäuresequenzen von 92,4% auf. Der hohe DDX3(X/Y) Homologiewert der AS-Sequenz ist, innerhalb der X-Y Genpaare der evolutionär jüngeren Strata (siehe Kapitel 1.6.), keine Ausnahme. Die Genpaare *USP9X/USP9Y* (91%), *UTX/UTY* (86%) und *ZFX/ZFY* (93%) weisen ebenfalls eine hohe Konserviertheit auf (Lahn & Page, 1999). Das evolutionär alte Genpaar *SOX3/SRY* (Stratum 1), welches nur innerhalb der charakteristischen HMG-Box homologisierbar ist, und die Amplikongene *CDY* (63%) und *RBMY* (61%) weisen dagegen deutlich geringe Homologien zu ihren X-Homologen auf.
Trotz der fehlenden Rekombination der Geschlechtschromosomen, existiert für die kodierenden *DDX3X* und *DDX3Y* Sequenzen also keine relaxierte Evolution. In Zellkulturexperimenten sind die zwei Humanproteine DDX3X und DDX3Y dementsprechend funktionell austauschbar und die Translationsfunktion der DDX3(X/Y) Proteine scheint, von Hefe bis Mensch, konserviert zu sein (Chuang et al., 1997; Fukumura et al., 2003, Sekiguchi et al., 2004). Im männlichen Geschlecht sind die Humanproteine DDX3X und DDX3Y aber funktionell nicht redundant.

Männer mit einer Y-chromosomalen AZFa-Deletion haben eine Keimzellpathologie namens „Sertoli Cell-only“ Syndrom, die zu einem Totalverlust der männlichen Keimzellen und Unfruchtbarkeit führt (Ditton et al., 2004; Vogt et al., 2008). Als Ursache des unterschiedlichen Protein-Expressionsmusters konnte, in dieser Arbeit die deutlich divergente Sequenzevolution der stromaufwärts gelegenen *cis*-regulativen Elemente identifiziert werden.

Die evolutionäre Betrachtung der *DDX3X* und *DDX3Y* Promotersequenzen ist ein Beispiel, wie die Allel-spezifische Sequenzevolution sowohl zu einer innerartlich differenzierten, als auch speziesspezifischen Evolution der *DDX3X* und *DDX3Y* Genregulation führte.

5.1. Konserviertheit und spezifische Evolution der *DDX3X* und *DDX3Y* Core-Promotermotive innerhalb der Säuger

Der in dieser Arbeit durchgeführte Vergleich der *DDX3X* und *DDX3Y* Transkriptionsstartstellen und Core-Promotermotive weist daraufhin, daß in den Säugetierordnungen der Primaten (Mensch, Schimpanse, Rhesusaffe, Weißbüschelaffe), der Rodentia (Maus) und der Artiodactyla (Rind) sowohl konservierte, als auch speziesspezifische Regulationsmodule für die Expression beider Gene in den verschiedenen Spezies, etabliert wurden. Die große Homologie der *DDX3Y* und *DDX3X* Core-Promotermodule in der Haplorrhini-Evolution fällt auf. Liang et al. (2008) berichten, daß Core-Promoterregionen, innerhalb der Primaten, im Vergleich zu kodierenden Genabschnitten, einer schnelleren Veränderung unterliegen und so zu artspezifischen Genexpressionsmustern führen. Dies ist für *DDX3X*, wie für *DDX3Y* nicht der Fall. Eine in Primaten konservierte *DDX3X* und *DDX3Y* Transkriptionskontrolle ist somit für beide Gene zu erwarten.

Die Sicht auf den Core-Promoter hat sich in den vergangenen Jahrzehnten verändert und gilt heute nicht nur als passive Montageplattform des Präinitiationskomplexes, sondern als wichtigstes Genregulationselement der Transkriptionskontrolle auf der DNA (Sverdlov & Vinogradova, 2010). Die vergleichende Analyse der Core-Promotermotive liefert daher wichtige Informationen über die Transkriptionsregulation von *DDX3X* und *DDX3Y* in den verschiedenen Spezies. Ausgehend von *DDX3* als gemeinsamen Ursprung von *DDX3X* und *DDX3Y* vor ~ 130 Mio. Jahren, weisen die drei Human *DDX3X* und vier *DDX3Y* Core-Promoterregionen eine deutlich komplexe Struktur auf. Eine Allel-spezifische Kontrolle der Genexpression von *DDX3X* und *DDX3Y* scheint die Konsequenz der Sequenzevolution.
Der hier beschriebene Effekt der Konserviertheit der AS-Sequenzen zwischen einem X-Y homologen Genpaar und Veränderung der regulativen Sequenzen, ist nicht exklusiv für *DDX3X* und *DDX3Y*. Das Human Genpaar *RPS4X/RPS4Y* (Ribosomal Protein S4) zeigt vergleichbare

Eigenschaften (Bergen et al., 1998; Andrés et al., 2008; Lopes et al., 2010). Das auf dem Y-Chromosom duplizierte *RPS4Y2* zeigt zudem eine hodenspezifische Expression, was wahrscheinlich durch Veränderung der Promoter und Core-Promoterregion ermöglicht wurde (Andrés et al., 2008; Lopes et al., 2010). Das Beispiel der Gene *CHD1Z* und *CHD1W* (Chromodomain-Helicase-DNA binding Protein) beim Zebrafinken (*Taeniopygia guttata*), einem Singvogel, zeigt, daß der Mechanismus der Promoterdivergenz nicht nur im Säuger-Gonosomensystem (♀ XX, ♂ XY), sondern auch im Vogel-System (♀ ZW, ♂ ZZ) möglich ist (Agate et al., 2004). An der Maus Protamin 2 Promoterregion wird zudem deutlich, daß wichtige regulative Sequenzen konservierter Proteine, auch in nah verwandten Mausspezies, in evolutionär kurzen Zeiträumen, speziesspezifisch verändert werden können (Martin-Coello et al., 2009).

Die mRNA-Expressionsergebnisse dieser Arbeit wurden mit konventionellen Endpunkt-RT-PCRs ermittelt. Eine Quantifizierung der mRNAs ist mit dieser RT-PCR-Methode nicht möglich, da eine optimale, exponentielle Amplifikation nicht kontinuierlich gewährleistet ist und die Endpunkt-Analyse nach 35 PCR-Zyklen im Agarosegel somit zu falschen quantitativen Beobachtungen führen kann. Ob es sich bei beobachteten Unterschieden in der Intensität der PCR-Produktbanden, in der Agarosegelelektrophorese, also um tatsächliche Unterschied der Transkriptmengen zwischen den Spezies handelt, z.B. durch kleinste Divergenzen in regulativen Sequenzen, oder ob der beobachtete Unterschied technisch bedingt ist, z.B. durch Unterschiede der Sekundärstrukturen im Bereich der Primerbindungsstellen, war mit der durchgeführten qualitativen RT-PCR nicht zu klären. Als Beispiel seien die möglichen Speziesunterschiede der *DDX3X* Genaktivität genannt. *DDX3X* Transkripte mit Start in T-TSS-III weisen scheinbar quantitative Unterschiede auf (Abb. 4-12, 4-36, 4-37, 4-38). Im Unterschied zu Mensch und *Callithrix* konnten für Schimpanse und *Macaca*, mittels konventioneller Endpunkt-RT-PCR, nur geringe Transkriptmengen mit Primern direkt 3´ der *DDX3X* T-TSS-III Position, innerhalb der Pyrimidin-reichen Region, identifiziert werden. Für den Javaneraffen *Macaca fascicularis*, einer nah verwandten Spezies zum Rhesusaffen, ist eine 5´ vollständige, hodenspezifische cDNA bekannt (GenBank Acc.No. AB168192.1), die in derselben *DDX3X* T-TSS-III Position startet. Dies weist daraufhin, daß auch in *Macaca mulatta DDX3X* T-TSS-III ein echter Transkriptstart sein sollte.

Angemerkt werden soll auch, daß mit der sehr sensitiven RT-PCR-Methode bereits geringste mRNA-Mengen amplifiziert werden können. Chelly et al. (1989) konnten zeigen, daß eigentlich gewebespezifisch exprimierte Transkripte, in geringsten Mengen auch ektopisch in anderen Geweben zu identifizieren sind, obwohl diese dort nicht zur Synthese eines Proteins führen. Generell wird davon ausgegangen, daß für eine effiziente Translation eine bestimmte, funktionelle mRNA-Menge benötigt wird. Ditton et al. (2004) konnten *DDX3Y* Transkripte, in

geringer Quantität in verschiedenen somatischen Humangeweben per RNA-Blot nachweisen. Zur Translation gelangen die *DDX3Y* Transkripte aber nur in der männlichen Keimbahn. Ob die RT-PCR-Produkte, bei 35 PCR-Zyklen, einer funktionellen mRNA-Menge entsprechen, oder eher einer ektopischen nicht-funktionellen Hintergrundtranskription zuzuschreiben sind, läßt sich nicht klären.

Eine mögliche Erklärung für die beobachteten Expressionsunterschiede zwischen den Spezies könnten auch Unterschiede im Histon-Methylierungszustand der Core-Promoterregionen sein (Cain et al., 2011). Die H3K4me3 Trimethylierung im Bereich der TSS ist als positives Transkriptionsinitiationssignal auf Chromatinebene bekannt (Li et al., 2007).

In lymphoblastoiden Zellinien von Mensch, Schimpanse und *Macaca* konnten speziesspezifische Unterschiede im H3K4 Trimethylierungsgrad identifiziert werden, die mit speziesspezifischen quantitativen Expressionsunterschieden der jeweiligen Gene korrelierten (Cain et al., 2011). Ebenfalls besteht die Möglichkeit Transkripte mit unproduktivem oder fehlerhaftem Splicing, mittels der sensitiven RT-PCR-Methode, zu erfassen (Sorek et al., 2004). Hodenspezifische Rind *Ddx3y* Transkripte mit Start in der Rind *MSY2* Sequenz könnten ein Beispiel für ein solches fehlerhaftes Splicing sein. Die Transkriptmengen waren teilweise noch nach 35 PCR-Zyklen unterhalb der Nachweisgrenze in der Agarosegelelektrophorese und erst nach Klonierung eines PCR-Amplifikatgemisches zu detektieren. Selbst wenn diese Transkripte weder funktionell sind, noch einer funktionellen mRNA-Menge entsprechen, zeigt es doch, daß noch weitere mRNA-Spezies vorhanden sein können, die mit den verwendeten Techniken nicht direkt zu detektieren sind.

5.1.1. *DDX3X* Transkriptionsstartstellen

TSS-I:

Die experimentellen Daten dieser Arbeit legen nahe, daß die Human *DDX3X* Genexpression von mindestens drei Core-Promoterregionen kontrolliert wird. Die proximale *DDX3X* TSS-I (BAC-Klon RP13-13A3, Pos. 30.639) mit einem fokussierten, dreiteiligen Core-Promotermodul (TATA-BRE^{d}-Inr) liegt in einem CpG-reichen DNA-Abschnitt (Abb. 4-9, 4-47). Lim et al. (2004) zeigten experimentell, daß dreiteilige Core-Promotermodule aus der Motivkombination TATA-Inr-MTE eine wesentlich höhere Transkriptionsaktivität zeigen, als zweiteilige Core-Promotermodule aus TATA-Inr oder Inr-MTE. Jin et al. (2006) konnten daraufhin mit statistischen Computermodellen zeigen, daß zwei dreiteilige Core-Promotermodule (BRE^{u}-Inr-MTE und TATA-Inr-MTE) in Human Promoterregionen tatsächlich favorisiert werden. Als mögliche Begründung gilt die mittige Lage der TSS zwischen zwei außen liegenden Motiven, was eine besonders stabile Positionierung des PIC gewährleisten soll (Jin et al., 2006). Die vorliegende *DDX3X* TSS-I Motivkombination (TATA-BRE^{d}-Inr) ist in der Literatur bisher funktionell nicht beschrieben. Deng & Roberts (2006) weisen aber darauf hin, daß BRE^{d}-Motive

die Promoteraktivität positiv modulieren können. Die Hypothese, daß es sich bei dem dreiteiligen *DDX3X* TSS-I Modul ebenfalls um ein effizientes Core-Promotermodul handelt, ist also vertretbar. Die Funktionalität wird durch den hohen Grad an evolutionärer Konserviertheit dieses *DDX3X* TSS-I Core-Promotermoduls in zehn Säugerspezies (von Opossum bis Mensch) unterstrichen (Abb. A-4). Dabei fällt auf, daß nicht nur die Motive konserviert sind, sondern auch die funktionell wichtigen Abstände zwischen den einzelnen Motiven. Dieses Regulationsmodul für die *DDX3X* Transkription hat sich somit seit ~ 150 Mio. Jahren, seit Trennung der Metatheria und Eutheria, kaum verändert (Abb. A-4). Die in dieser Arbeit durchgeführten RT-PCR Analysen zur Kartierung der *DDX3X* TSS-I haben gezeigt, daß diese Region in allen untersuchten Spezies auch für die Transkriptionsinitiation genutzt wird und nicht nur eine Sequenzhomologie im Alignment aufweist (Abb. 4-10, 4-33, 4-34, 4-35, 4-43, 4-44, A-4).

Die hochkonservierte *DDX3X* TSS-I Region wird durch die Daten von Waterston et al. (2002) und Jin et al. (2006) bestätigt, daß in funktionell orthologen Mensch-Maus Genpaaren *cis*-regulative Elemente konserviert bleiben können. Sobald allerdings keine Funktion mehr auf bestimmten *cis*-regulative Elemente liegt, degenerieren diese, wie die Beispiele von *DDX3X* T/B-TSS-II und T-TSS-III zeigen, die in Maus nur noch mit reduzierter Homologie zu finden sind und keine Transkriptionsaktivität mehr besitzen (Abb. 4-43, Abb. A-4). Dies stimmt wiederum mit der Feststellung von Mayr (2005) und Odom et al. (2007) überein, daß in den Abstammungslinien der Nagetiere und Primaten eine erhöhte Evolutionsgeschwindigkeit (Selektion) existiert, die vor allem für gewebespezifisch exprimierte Gene artspezifische Expressionsmuster etabliert.

T/B-TSS-II:

Für die fokussierte Human *DDX3X* T/B-TSS-II Position (BAC-Klon RP13-13A3, Pos. 30.314), mit einem dreiteiligen Core-Promotermodul (TATA-Inr-DPE) konnte im männlichen Geschlecht eine Gewebespezifität für Hirn und Hoden festgestellt werden (Abb. 4-11). Die *DDX3X* T/B-TSS-II ist auf Grund der distinkten Position und der Gewebespezifität als alternativer *DDX3X* Core-Promoter zu definieren (Baek et al., 2007; Davuluri et al., 2008; Jacox et al., 2010; Sverdlov & Vinogradova, 2010). Das dreiteilige *DDX3X* T/B-TSS-II Core-Promotermodul ist nur innerhalb der Primaten konserviert (Abb. A-4). Allerdings ist mit abnehmender Homologie das dreiteilige Modul auch in Rind, Maus und Opossum zu identifizieren (Abb. A-4). Dabei fällt auf, daß das DPE-Motiv in allen zehn Spezies perfekt konserviert ist. Die Transkriptinitiation bei T/B-TSS-II wurde experimentell für Mensch und Rind anhand spezifischer RT-PCR Analysen gezeigt (Abb. 4-11, 4-44). Es ist anzunehmen, daß in den anderen Primatenspezies ebenfalls eine Initiation in T/B-TSS-II möglich ist.

Interessant ist, daß bereits in Rind eine Gewebespezifität für diese Initiationsstelle festzustellen ist. Sie wurde somit bereits vor ~ 100 Mio. Jahren etabliert. Obwohl in der Maus das DPE-Motiv

ebenfalls perfekt konserviert vorliegt und im korrekten Abstand ein Y-R Dinukleotid festzustellen ist (Abb. A-4), konnte für die Maus keine Initiation in diesem Bereich festgestellt werden (Abb. 4-43). Dieser speziesspezifische Maus-Mensch Unterschied korrespondiert mit den Daten von Odom et al. (2007), die zeigen, daß die Genexpression zwischen Mensch und Maus besonders in Bezug auf die gewebespezifische Transkriptionskontrolle deutlich verschieden ist.
Auf Grund der perfekten Konserviertheit des DPE-Motives ist anzunehmen, daß es sich bei diesem Core-Promotermodul um einen DPE-abhängigen Core-Promoter handelt und die Anwesenheit des DPE-Motives funktionell wichtig für die gewebespezifische Expression ist. Die in *DDX3X* T/B-TSS-II realisierte Motivkombination (TATA-Inr-DPE) ist bekannt, kommt aber wesentlich seltener vor, als zweiteilige Inr-DPE Core-Promotermodule (Juven-Gershon & Kadonaga, 2010). Eine Gewebespezifität könnte über den DPE-binde Faktor TAF9 (Expression in Hodengewebe, HPA), bzw. den paralogen Faktor TAF9b, generiert werden (Frontini et al., 2005; Sengupta et al., 2009). Die TAF9-Faktoren können mit TFs interagieren und so Spezifitäten erzeugen (Sengupta et al., 2009). Auch ist noch nicht klar, ob TAF9 und TAF9b die identischen TF- und DNA-Bindeeigenschaften besitzen (Frontini et al., 2005). Juven-Gershon et al. (2008b) konnten zudem zeigen, daß das *Drosophila* Gen *caudal* bevorzugt DPE-abhängige Gene, vor allem verschiedenen Hox Gene, aktiviert. Dies spricht dafür, daß eine bestimmte Verbindung zwischen den Aktivator-TFs, die an die Promoter-DNA binden, und den Core-Promotermotiven besteht. Die homologen Vertebraten Caudal-Proteine (Cdx1, Cdx2, Cdx4) sind ebenfalls in der Expressionskontrolle von Hox Genen beteiligt (Juven-Gershon et al., 2008b).

T-TSS-III:

Die Human *DDX3X* T-TSS-III (BAC-Klon RP13-13A3, Pos. 29.825) ist ebenfalls ein fokussierter, alternativer *DDX3X* Core-Promoter mit dreiteiligem Modul (TATA-Inr-DPE) und mit Spezifität in Hodengewebe (Abb. 4-12). Vergleichbar zum Maus *Dmrt7* Gen (Kawamata et al., 2007), konnte für das X-chromosomale Humangen *DDX3X* nicht nur eine differentielle Expression zwischen Keimzellen und Soma im männlichen Geschlecht festgestellt werden, sondern auch ein geschlechtsspezifischer Unterschied, mit Transkriptinitiation in der *DDX3X* T-TSS-III Position ausschließlich in männlichen Keimzellen. Im Hodentubulus konnte die Keimzellspezifität der *DDX3X* Transkripte, mit Start in T-TSS-III, in ersten RNA *in-situ* Hybridisierungen gezeigt werden (Jakob, 2005). Eine Expression der *DDX3X* Transkripte mit Start in T-TSS-III konnte in weiblichen Geweben (Ovar, Niere, Leukozyten) nicht gezeigt werden (Zimmer J., unveröffentlichte Resultate). Die Transkription von *DDX3X* ist somit in Human geschlechts- und gewebespezifisch regulierbar.
Innerhalb der Primaten (Mausmaki bis Mensch) besteht eine hohe Konserviertheit für das *DDX3X* T-TSS-III Core-Promotermodul, wiederum mit der höchsten Homologie für das DPE-

Motiv (Abb. A-4). In Maus und Rind sind positionshomologe Inr- und DPE-Motiv noch erkennbar. Dagegen ist im Opossum für den positionshomologen Bereich keine Homologie der Core-Promotermotive erkennbar (Abb. A-4). Eine Expression langer hodenspezifischer *DDX3X* Transkripte konnte per RT-PCR Analyse nur für die vier untersuchten Primatenspezies (Human, Schimpanse, *Macaca*, *Callithrix*) festgestellt werden.
Somit ist festzuhalten, daß die *DDX3X* Promoterregion über drei klar fokussierte Core-Promotermodule, in einer langen (C,G)-reichen Sequenzdomäne, charakterisiert ist (Abb. 4-9). Dies stimmt mit Rach et al. (2011) überein, die beschreiben, daß fokussierte Core-Promoter meist durch bestimmte *cis*-Elemente, wie Core-Promotermotive und assoziierte TFBS definiert werden. Die *DDX3X* Promoterregion fällt in diesem Punkt auf, da in CpG-reichen Promoter-regionen der breite Core-Promoter, oder Initiatorregionen ohne weitere Core-Promotermotive, häufig die vorherrschende Core-Promoterform darstellen (Juven-Gershon & Kadonaga, 2010).

Das Vorhandensein zwei aktiver alternativer *DDX3X* Core-Promoter, zusätzlich zum konstitutiven TSS-I Core-Promoter, ist im Einklang mit den aktuellen Daten, daß ~ 50% aller kodierenden Humangene über einen oder mehrere alternative Core-Promoter verfügen, die dadurch die Transkriptdiversität des Humangenoms enorm steigern können (Kimura et al., 2006; Ma et al., 2009). Alternative Keimzell-spezifische Genexpressionen sind z.B. für die Mausgene *Gsg2* (Haspin) und *mGstm5* bekannt (Tokuhiro et al., 2007; Dehari et al., 2008).

Bei Sequenzabweichungen zu den publizierten Core-Promoter Konsensussequenzen (siehe Juven-Gershon & Kadonaga, 2010) ist häufig nicht klar, oder nicht bekannt, ob diese Mutationen zu einer Verstärkung, Erniedrigung oder sogar zu einem Totalausfall der Genexpression führen können (Savinkova et al., 2009; de Vooght et al., 2009). Selbst identische TATA-Mutationen können je nach Promoterkontext zu unterschiedlichen Effekten führen (Savinkova et al., 2009). In der dbSNP (NCBI-Datenbank für kurze genetische Varianten) konnte für keines der *DDX3X* und *DDX3Y* Core-Promotermodule ein Sequenz-Polymorphismus identifiziert werden.
Das potentielle BRE^{d} mit der Sequenz „ATCCGTG“ im Human *DDX3X* TSS-I Core-Promoter ist ein Beispiel für eine Abweichung zur Konsensus-sequenz „RTDKKKK“ in zwei Nukleotiden. Die zwei nicht-kanonischen Cytosin-Nukleotide sind aber von Opossum bis Mensch konserviert (Abb. A-4). Lediglich in der Maus ist das zweite Cytosin durch ein kanonisches Guanin ersetzt. Da diese zwei nicht-kanonischen Nukleotide in der Säugerevolution stabilisiert wurden, stellt sich die Frage, ob es sich hierbei tatsächlich um eine negative Mutationen in dem Core-Promotermotiv handelt. Eine negative Mutation sollte der Theorie nach eigentlich durch die reinigende Selektion wieder eliminiert werden. Diese Nukleotidbasen sind aber wie eine positive, oder neutrale Mutation stabilisiert worden. Ebenfalls nicht bekannt ist, wie sich

zusätzliche artspezifische Unterschiede in der Sequenz auf die Transkriptionsaktivität auswirken. Im Schimpansen befindet sich an der ersten BRE^{d}-Position ein nicht-kanonisches Cytosin anstatt des konservierten Purins (Abb. A-4).
Die Bestimmung der BRE^{d}-Konsensussequenz ging ursprünglich von der Sequenz des Adenovirus E4 (AdE4) Promoters aus, für die ein Kontakt des GTF TFIIB an einen Bereich stromabwärts der TATA-Box gezeigt werden konnte (Deng & Roberts, 2005). Die original AdE4 BRE^{d}-Sequenz (CGCTCTG) weist vier Abweichungen zur später bestimmten Konsensus-sequenz auf und zeigt dennoch eine effektive Bindung zu TFIIB (Deng & Roberts, 2005). Der verwendete selektive Bandshift-Assay hat auch gezeigt, daß Cytosine an Position drei und vier der BRE^{d}-Sequenz vorkommen können (Deng & Roberts, 2005). Da dies in den selektionierten Sequenzen aber seltener vorkam, wurden die Cytosine nicht als kanonische Nukleotide für die Positionen drei und vier in der BRE^{d}-Konsensussequenz geführt (Deng & Roberts, 2005).

Das *DDX3X* 5´UTR-Splicing, in Transkripten mit Start in T-TSS-III, konnte nur in Human und nur im Hodengewebe identifiziert werden, obwohl das gleiche GC-AG Splicesignal in allen Haplorrhini konserviert vorliegt (Abb. 4-32, A-4). Vergleichbar dazu existiert eine ebenfalls hodenspezifische Human *HEMGN* Transkriptvariante (*HEMGN-t*), die sich in 5´- und 3´UTR-Länge, nicht aber im ORF, von der hämatopoetischen Variante (*HEMGN-h*) unterscheidet (Yang et al., 2003). Auch das *HEMGN* Gen addiert ein hodenspezifisches 5´UTR Exon-1t über ein GC-AG Splicesignal. Ebenfalls vergleichbar zu den hodenspezifischen *DDX3X* Transkripten, wird die *HEMGN* 3´UTR in Hodengewebe an einem internen, proximalen Polyadenylierungssignal prozessiert, um eine verkürzte 3´UTR zu erzeugen (Yang et al., 2003). In Maus konnte die *Hemgen-t* Transkriptvariante, per RNA *in-situ* Hybridisierungen, in runden Spermatiden bestätigt werden (Yang et al., 2003).
Es ist bekannt, daß speziesspezifische Unterschiede in der Aktivität alternativer Splicemuster, trotz Sequenzhomologie, existieren (Skandalis et al., 2010). Da bis zu 50% der in T-TSS-III startenden Human *DDX3X* Transkripte in der 5´UTR gesplict vorkommen können (Zimmer J., unveröffentlichte Resultate), wird ein artifizielles Splicing ausgeschlossen. Für Eukaryoten, z.B. bei *Drosophila* ist bekannt, daß vor allem im Hodengewebe alternative Splicemechanismen existieren, die zu unterschiedlichen Transkriptvarianten mit unterschiedlichen Translations-eigenschaften führen können (Taliaferro et al., 2011).

5.1.2. *DDX3Y* Transkriptionsstartstellen

TSS-I:

Die Human *DDX3Y* Genexpression wird von vier Core-Promoterregionen kontrolliert, die sich deutlich von den drei Human *DDX3X* Core-Promotermodulen unterscheiden. Im Gegensatz zum hochkonservierten, fokussierten *DDX3X* TSS-I Core-Promotermodul, ist für die ubiquitäre

DDX3Y Expression proximal ein breiter TSS-I Core-Promoter (BAC-Klon 475I1, Pos. 55.032 - 55.104) entstanden (Abb. 4-2). Im Gegensatz zu einem fokussierten Core-Promoter werden für breite Core-Promoterregionen meist schwächere Transkriptionsinitiationspotentiale beschrieben (Sandelin et al., 2007; Juven-Gershon & Kadonaga, 2008). Das gesamte Initiationspotential verteilt sich dabei auf mehrere TSS, die je nach Gewebe auch unterschiedlich aktiv sein können, so daß jede einzelne TSS nur ein geringes Initiationspotential besitzt. Computermodelle, anhand von genomweiten TSS- und Chromatinkarten, unterstützen die These, daß die Chromatineigenschaften bei breiten Core-Promotern eine größere Rolle für die Initiation spielen, als bei einem fokussierten Core-Promoter (Rach et al., 2011). Die Definition von zwei verschiedenen Core-Promotertypen spiegelt auch die Aussage von Juven-Gershon & Kadonaga (2010) wieder, nach der sich Transkriptionsinitiationsmechanismen an fokussierten und breiten Core-Promotern grundlegend unterscheiden.

Die breite *DDX3Y* TSS-I Core-Promoterregion umfaßt 73 bp und wird von zwei perfekten Initiator-Motiven (TSS-Ia, TSS-Id) eingegrenzt, für die jeweils kein zweites assoziiertes Motiv identifiziert wurde. Die zwei in der DBTSS-Datensammlung prädominanten Initiationsstellen TSS-Ib (Inr-DPE) und TSS-Ic (TATA-Inr) sind jeweils zweiteilig aufgebaut (Tab. 4-1; Abb. 4-2).

Der hier beschriebene Aufbau der *DDX3Y* TSS-I Region konnte nur innerhalb der Haplorrhini konserviert gefunden werden. Für Maus und Rind konnte per RT-PCR ebenfalls eine positionshomologe *Ddx3y* TSS-I Region mit wahrscheinlich ubiquitärer Expression identifiziert werden. Y-R Dinukleotide konnten *in-silico* für beide Spezies identifiziert werden. Auf Grund der erheblichen Sequenzunterschiede scheint es allerdings unwahrscheinlich, daß es sich um konservierte positionshomologe Core-Promotermotive handelt.

TSS-I-(T):

Das präferentiell in Hodengewebe aktive Core-Promotermodul (Inr-DPE), mit Transkriptionsstart in TSS-I-(T) (BAC-Klon 475I1, Pos. 54.911), ist auch in den untersuchten Primaten (Schimpanse, *Macaca*, *Callithrix*) sowie in Rind zu finden. Da diese TSS 121 bp stromaufwärts der Human TSS-Ia Position lokalisiert ist und hauptsächlich in Hodengewebe aktiv ist, wurde die *DDX3Y* TSS-I-(T) als eigenständige, alternative TSS definiert. Die *DDX3Y* TSS-I-(T) ist Y-spezifisch und hat kein homologes Modul in der *DDX3X* Promoterdomäne.

T-TSS-I:

Die hodenspezifisch genutzte *DDX3Y* T-TSS-I in BAC-Position 54.317 (BAC-Klon 475I1), ist Bestandteil eines dreiteiligen Core-Promotermoduls (TATA-Inr-MTE). Diese erste Initiationsstelle in Exon-T ist deutlich fokussiert und in den untersuchten Primaten konserviert (Abb. 4-15; Abb. 4-26). Wahrscheinlich ist auch, daß hodenspezifische *Ddx3y* Rindertranskripte mit Exon-T in der homologen Position initiieren können (Abb. 4-27; 4-29). Trifft dies zu, so wäre

die T-TSS-I eine seit ~ 100 Mio. Jahren konservierte TSS für die hodenspezifische Expression von *DDX3Y* Transkripten in Keimzellen. Das innerhalb der Primaten und Rind funktionell konservierte Exon-Ta 5´GT-Splicesignal deutet ebenfalls darauf hin, daß Keimzell-spezifische *DDX3Y* Transkripte mit dem zusätzlichen 5´UTR Exon-T bereits vor der Trennung der Abstammungslinien von Rind und Primaten vor ~ 100 Mio. Jahren existierten. Das Exon1b 3´AG-Splicesignal ist innerhalb der Primaten funktionell konserviert. In Rind konnte die Nutzung eines weiteren 3´AG-Splicesignals, Exon1c genannt, festgestellt werden (Abb. 4-27). Die Entstehung weiterer *DDX3Y* 5´UTR-Splicevarianten (Ta-1a, Tb-1b) in Human, korreliert mit der Feststellung, daß alternative Splicemuster, trotz Sequenzhomologie in verwandten Spezies, häufig speziesspezifisch etabliert werden (Skandalis, et al., 2010). Interessanterweise wird in der Primaten *DDX3Y* Exon-Ta-1b Splicevariante der komplette TSS-I Bereich aus dem Transkript eliminiert, indem direkt vor das Translationsstartcodon ATG gesplict wird.

In der Maus ist ein weitreichender Sequenz- und Strukturumbau der positionshomologen *Ddx3y* 5´Region, ohne erkennbare Homologie des T-TSS-I Core-Promotermoduls und ohne Erhalt von Exon-T erfolgt (Abb. 4-27). Wahrscheinlich ist bei der Maus der evolutionäre Druck auf den funktionellen Erhalt der *Ddx3y* Keimbahnfunktion, nach Entstehung des autosomalen Retrogens *D1Pas1*, aus *Ddx3* Transkripten, verloren gegangen (Mazeyrat et al., 1998; Session et al., 2001; Vong et al., 2006). Ein entscheidender Faktor für die größere Sequenzdivergenz in der Maus, im Vergleich zu Mensch und Rind, dürfte die wesentlich höhere Generationszahl bei gleicher Evolutionszeit sein.

T-TSS-II:

Die Keimzell-spezifische *DDX3Y* T-TSS-II, mit zweiteiligem Core-Promotermodul (TATA-Inr), liegt in einem (A,T)-reichen *DDX3Y* Promoterabschnitt (Abb. 4-2). Diese Initiationsstelle ist besonders auffällig, da sie in einer repetitiven Minisatellitensequenz, *MSY2* genannt, entstanden ist (Rauschendorf et al., 2011). Die *DDX3Y* T-TSS-II Position (BAC-Klon 475I1, Pos. 54.051) konnte im 3´ Bereich der zweiten von vier *MSY2* Repeateinheit kartiert werden und stellt die am weitesten 5´ gelegene Position dar, die mit der konventionellen RT-PCR Methode gesichert kartierbar war.

Das *DDX3Y* T-TSS-II Core-Promotermodul mit TSS in BAC-Position 54.051 scheint auf Grund des kanonischen TATA-Motives „TTATAATC“ in *MSY2-2* genutzt zu werden (Abb. 4-4). Eine Initiation an einem der vielen weiteren potentiellen TATA- und Inr-Motive, die durch die (A,T)-reiche *MSY2* Repeatsequenz begünstigt werden, ist aber durchaus vorstellbar. So sind die humanspezifischen Core-Promotermodule am Übergang von Repeat 2 zu 3 und 3 zu 4 mit pot. TSS in den BAC-Positionen 54.065 (*MSY2-3*) bzw. 54.164 (*MSY2-4*), ebenfalls geeignete

Kandidaten für die gesuchte *DDX3Y* TSS in *MSY2* (Abb. 4-3). Wenn eine Initiation von *DDX3Y* Transkripten im 5´ Bereich des dritten Repeats möglich ist, sollte dieses auch für den 5´ Bereich des vierten Repeats gelten, da die beiden Repeateinheit zu 100% homolog sind. Für eine hodenspezifische Initiation im 3´ Bereich von Repeat 1 bzw. in der 5´ Region von Repeat 2, stromabwärts des optimalen TATA-Motives „TATATAAT“ in *MSY2-1*, gibt es im Unterschied zu Schimpanse und *Macaca*, für Human bisher keine experimentelle Evidenz (Rauschendorf et al., 2011). Die Transkriptionsinitiation im verdoppelten, bzw. vervierfachten *MSY2* Repeat scheint eine Spezifität der catarrhinen Primaten (Mensch, Schimpanse, *Macaca*) zu sein, da für *Callithrix* keine RNA-Expression innerhalb des *MSY2* Repeats festgestellt werden konnte (Abb. 4-26).
Die generelle Tendenz der (A,T)-reichen *MSY2* Sequenz zur Bildung von pot. Core-Promoter-motiven in verschiedenen Säugerspezies scheint allerdings konserviert zu sein, auch wenn die beobachtete Initiation von Rind *Ddx3y* Transkripten im Bereich von *MSY2* (Abb. 4-29, 4-30), wahrscheinlich nicht in identischen Motiven erfolgt.

Generell wird davon ausgegangen, daß Minisatellitensequenzen in verwandten Spezies eine Hypervariabilität durch ihre Keimbahn-Instabilität aufweisen (Williams et al., 2001; Bois & Jeffreys, 1999; Jeffreys et al., 1998). Die in Primaten identifizierte hohe Konserviertheit der *MSY2* Repeats, zwischen 78 und 100%, deutet somit darauf hin, daß diese Sequenzen einer funktionellen Selektion unterliegen und nicht hypervariabel sind.
Ein Minisatellit im Intron9 des Humangens „Glucosephosphat-Isomerase“ (*GPI*) zeigt ebenfalls eine hohe Konserviertheit in mehreren Säugerspezies (Williams et al., 2001). Reportergen-experimente mit diesem Minisatelliten haben aktivitätssteigernde Eigenschaften auf SV40-, CMV- und Human GPI-Promoter gezeigt (Williams et al., 2001).
Minisatelliten und allgemein repetitive Sequenzen sind dafür bekannt, daß sie Einfluß auf die Transkription benachbarter Gene haben (Nakamura et al., 1998; Vergnaud & Denoeud, 2000; Iwashita et al., 2001; Jacox et al., 2010). Beispiele hierfür sind *GPI*, das Humangen für den Serotonintransporter *5-HTT*, das Human Dopamintransportergen *SLC6A3*, das Humangen Secretin und das Human Insulingen (*INS*) (Kennedy et al., 1995; Michelhaugh et al., 2001; Williams et al., 2001; Inoue-Murayama et al., 2008; Lee et al., 2008). Auch ist bekannt, daß Minisatellitensequenzen häufig Purin-reiche (A,G) Abschnitte enthalten, die zu nicht-kanonischen DNA-Formen führen können (Mirkin, 2007). Dieses kann wiederum zu einem erleichterten Zugang des Core-Promoter-Erkennungskomplex zur DNA führen. Die 409 bp lange Human *MSY2* Minisatellitensequenz ist sowohl (A,T)-, als auch (C,T)-reich. Lokal kann dies zu unterschiedlichen Nukleotidverteilungen führen. Der 5´Bereich der *MSY2* Repeateinheiten 1 und 2, der nicht in Repeat 3 und 4 vorkommt, ist z.B. Purin-reich (siehe *MSY2* Alignment in Abb. 4-3).

5.2. Auffällige Nukleotidkompositionen und „Simple Sequence Repeats“ (SSR) im Bereich der *DDX3X* und *DDX3Y* Core-Promoterregionen

Baek et al. (2007) konnten in einem Maus-Mensch Core-Promotervergleich, anhand einer Einteilung in konstitutive vs. alternative Core-Promoter und CpG-reiche vs. CpG-arme Core-Promotersequenzen, unterschiedliche Genontologien und Funktionen für die regulierten Gene identifizieren. Die Nukleotidverteilungen nehmen Einfluß auf Position und Dichte der Nukleosomen und somit die Transkriptionsregulation (Schones et al., 2008; Cairns, 2009; Radman-Livaja & Rando, 2010). Darvish et al. (2011) zeigten, daß neben CpG-Islands auch „Purin-Islands“ und „Pyrimidin-Islands“ stromaufwärts (-120 bp bis +1 bp relativ zur TSS) zu einer Transkriptionseinheit Auswirkungen auf die Transkriptionskontrolle haben können.

5.2.1. Nukleotidkompositionen im Bereich der *DDX3X* Core-Promotermodule

Die *DDX3X* TSS-I liegt in einer auffällig (C,G)- und (A,G)-reichen Sequenzregion (C,G = 64,5%; A,G = 56,2%). Die *DDX3X* T/B-TSS-II Position weist ebenfalls eine auffällige Basenzusammensetzung 5´ der Initiationsstelle auf. Die insgesamt CpG-reiche *DDX3X* Promoterregion (Abb. 4-47) ist 5´ der T/B-TSS-II deutlich (A,G)-reich und liegt somit in einem „Purin-Island“. Für *DDX3X* T/B-TSS-II konnte ein Purin-Anteil von 71,9% in diesen 121 bp festgestellt werden (A,T = 57,0%). In Tab. 4-8 konnte auch gezeigt werden, daß 5´ zur T/B-TSS-II Position eine fast vollständig homopolymere Adeninsequenz lokalisiert liegt. Diese 25 bp lange Sequenz kann auch als Mikrosatellit (Mononukleotidrepeat) gesehen werden. Die Identifizierung solcher repetitiver Elemente im Bereich der Core-Promoter ist von funktioneller Bedeutung, da bereits bekannt ist, daß repetitive Sequenzen, wie Mikrosatelliten, STRs („short tandem repeats“) und Minisatelliten, sowie Transposons Einfluß auf die Transkriptionsregulation nehmen können und auch zur Entstehung neuer Core-Promoterregionen beitragen können (Li, 1997; Jeffreys et al., 1985; Ali & Wallace 1988; Litt & Luty, 1989; Beilina et al., 2004; Tassone et al., 2007; Lawson & Zhang, 2008; Jacox et al., 2010; Darvish et al., 2011).

Das dreiteilige *DDX3X* T-TSS-III Core-Promotermodul (TATA-Inr-DPE) ist am Übergang einer sehr (C,G)-reichen zu einer sehr (C,T)-reichen Sequenz im *DDX3X* Promoter lokalisiert (Abb. 4-9). Der (C,G)-Gehalt in den 5´ 121 bp beträgt 73,6%, der (C,T)-Gehalt 52,1%. Die Pyrimidin-reiche Region scheint sich innerhalb der Altweltaffen stabilisiert zu haben und, vor allem in der Abstammungslinie von Mensch und Schimpanse, auch expandiert zu sein (Abb. A-4). Ein positiver Effekt auf die Promoteraktivität ist für (C,T)-reiche Promotersequenzen bekannt (Dörflinger et al., 1999; Han & de Lanerolle, 2008; Darvish et al., 2011). Auch ist bekannt, daß die Core-Promoterregionen vieler Keimzell-spezifischer Gene (C,G)-reich, kleiner als 100 bp

und inaktiv in somatischen Geweben sind (Upadhyaya et al., 2002). Dies wird z.B. durch gezielte Demethylierung von CpG-Island Promotern in Keimzellen erreicht, was folglich eine veränderte Chromatinstruktur und alternative Initiationsmöglichkeiten zur Folge hat (Upadhyaya et al., 2002; Tokuhiro et al., 2007). Stromabwärts der *DDX3X* T-TSS-III Position ist am 3´Ende der Pyrimidin-reichen Region in Human und Schimpanse ein nicht-kodierender Pentanukleotidrepeat $(TCCCC)_3$ mit drei Wiederholungen lokalisiert (Abb. A-4). Dieser SSR unbekannter Funktion liegt im *DDX3X* Transkript 119 bp stromabwärts der 5´Capstruktur, was mit der Beobachtung von Lawson & Zhang (2008) korrespondiert, daß SSRs gehäuft am 5´Ende einer mRNA und nicht in Nähe zum kodierenden Bereich vorkommen.

5.2.2. Nukleotidkompositionen im Bereich der *DDX3Y* Core-Promotermodule

Es fiel auf, daß sich die Zusammensetzung der Nukleotide im Bereich der breiten *DDX3Y* TSS-I Core-Promoterregion von 5´ (TSS-Ia) nach 3´ (TSS-Id), von (C,G)- und Purin-reich zu einem ausgeglichenem Verhältnis der Nukleotide, verändert. Die 5´ 121 bp zu TSS-Ia haben einen (C,G)-Gehalt von 56,5% und einen (A,G)-Gehalt von 65,3%. Für TSS-Id beträgt der (C,G)-Gehalt 52,1% und für den (A,G)-Gehalt 50,4%. Es besteht hier die Möglichkeit, daß die verschiedenen *DDX3Y* TSS-I Initiationsstellen von unterschiedlichen PICs in unterschiedlichen Geweben, oder zu unterschiedlichen Entwicklungszeitpunkten genutzt werden. Kawaji et al. (2006) konnten eine unterschiedliche Gewebeverteilung für breite TSS-Cluster in Maus Core-Promoterregionen zeigen. Die Mechanismen, wie eine bestimmte TSS in einem Gewebe selektioniert wird und in einem anderen nicht, sind aber bisher noch nicht geklärt. Da breite Core-Promoterregionen häufig mit CpG-Islands assoziiert sind, wird eine differentielle CpG-Methylierung als eine der Ursachen für die dynamische TSS-Nutzung diskutiert (Kawaji et al., 2006).

Die Basenzusammensetzung 120 bp stromaufwärts des *DDX3Y* TSS-I-(T) Moduls ist wiederum Purin-reich (58,7%). Das (C,G)- zu (A,T)-Verhältnis ist dabei mit 50% für beide ausgeglichen. Die Keimzell-spezifisch genutzten *DDX3Y* T-TSS-I (57%) und T-TSS-II (58,7%) liegen jeweils in (A,T)-reichen genomischen *DDX3Y* Sequenzabschnitten (Abb. 4-2). Das (C,T)- zu (A,G)-Verhältnis ist dabei in beiden Regionen ausgeglichen.

Ein nicht-kodierender SSR mit der Sequenz $(TGG)_4$, also ein Trinukleotidrepeat (TNR), ist in der 5´UTR des *DDX3Y* TSS-I-(T) Transkript 50 Nukleotiden stromabwärts zur 5´Cap lokalisiert. Es ist bekannt, daß Trinukleotid SSRs mit vier Wiederholungen in gewebespezifisch exprimierten Genen häufiger auftreten, als in Housekeeping-Genen (Lawson & Zhang, 2008). Die Instabilität und Dynamik bezüglich der TNR-Repeatzahl ist mit verschiedenen Krankheiten, vor allem neurodegenerativen Erkrankungen, assoziiert. Die „Fragile X-Syndrom“ Pathologie (FXS, OMIM

#300624) z.B., entsteht durch die Expansion eines CGG-Repeats im *FMR1* Gen (Beilina et al., 2004; Oostra & Willemsen, 2009). Der pathologische Effekt von TNRs, durch Expansion oder Deletion, kann durch mehrere Mechanismen, je nach Lage des Repeats, im Promoter (z.B. *FMR2*), in Introns (z.B. *FXN* Intron1), in 5´UTRs (z.B. *FMR1*), in 3´UTRs (z.B. *DMPK*), oder im Leserahmen (*HTT* Exon1), ausgelöst werden (Cummings & Zoghbi, 2000; Orr & Zoghbi, 2007; Kozlowski et al., 2010; McMurray, 2010). Dazu gehören chromosomale Brüche (z.B. Mikrodeletion im paternalen Chromosom 14 [14q32.2]), aberrante Promoter-Methylierungen (z.B. *FMR1*, *FMR2*) an CpG-Dinukleotiden in TNRs, keine oder verminderte Translationsfähigkeit der Transkripte (*DMPK*) und veränderte Proteine (z.B. Huntingtin).

Der *DDX3Y* $(TGG)_4$-Repeat konnte in den drei Catarrhini (Mensch, Schimpanse, *Macaca*) in gleicher Länge konserviert nachgewiesen werden (Abb. 4-27). Die Stabilität dieses TNRs, seit mehr als 29 Mio. Jahren (siehe Abb.1-6), weist auf ein funktionelles *cis*-Element hin. Da *DDX3Y* Transkripte mit Start in TSS-I-(T) präferentiell in Hodengewebe nachgewiesen werden konnten (Abb. 4-1), stellt sich die Frage, ob veränderte TGG-Repeatzahlen oder TGG-Mutationen in Human existieren und wenn ja, ob diese dann mit einer reduzierten männlichen Fertilität des betroffenen Mannes assoziiert sind? In dem kartierten $(TGG)_4$-Repeat konnten TFBS für die u.a. in Hoden exprimierten TFs ZNF202 und ZNF263 lokalisiert werden (Tab. A-16), die sowohl als Transkriptionsrepressor und -Aktivator fungieren können (Monaco et al., 1998; Wagner et al., 2000; Frietze et al., 2010).

Für das letzte G in der vierten TGG-Einheit existiert ein „Single Nucleotide Polymorphism“ (SNP), mit den Allelen G und A, unter der Referenz-SNP-Nummer „rs13304111“ (Ning et al., 2001). Der SNP eliminiert die ZNF263 TFBS. Die klinische Relevanz dieses SNP ist bisher nicht bekannt.

Ein expandierender $(TGG)_4$-Repeat könnte auch zu neuen oder mehr TFBS führen, die potentiell die *DDX3Y* Genexpression beeinflussen. Tatsächlich gibt es in der Literatur deutliche Hinweise, daß ein Teil der Human TFBSs aus repetitiven Sequenzen, wie SSRs und transponierbaren Elementen (TEs), stammt (Polavarapu et al., 2008).

5.3. Konserviertheit und spezifische Evolution der *DDX3X* und *DDX3Y* Promoterregionen

Die Suche nach konservierten regulativen Sequenzen in den Human *DDX3X* und *DDX3Y* Promoterregionen hat, neben konservierten X-Y Sequenzblöcken (KRiH-1 bis -8 und CNSiH-1), ebenfalls deutlich gezeigt, daß diese Sequenzen erhebliche Chromosomen-spezifische Veränderungen, seit dem Fehlen der Gonosomenrekombination vor ~ 130 Mio. Jahren, erfahren haben (Abb. 4-47). Die Homologie in den Promoterregionen bis 2,5 kb stromaufwärts des konservierten DDX3(X/Y) Translationsstartcodons beträgt insgesamt 50% und zeigt keine

Konserviertheit der identifizierten Core-Promoterregionen. So überrascht es auch nicht, daß 24 der 30 kartierten, X-Y konservierten TFBS (in 18 unabhängigen Positionen) in den neun X-Y konservierten Sequenzblöcken konzentriert sind (Tab. 4-10). Die Mehrzahl (19 von 30) dieser X-Y konservierten Sequenzmotive sind potentielle Bindestellen für TFs, von denen bereits eine Expression in Hodengewebe bekannt ist (Abb. 4-48). Mit den zwei gewählten Ansätzen zur Identifizierung konservierter X-Y Bereiche, das mVISTA Sequenzalignment, sowie die Kartierung konservierter TFBS, konnte also eine gute Übereinstimmung festgestellt werden. Die zwei Analysen bestätigen somit unabhängig voneinander die kartierten X-Y Bereiche.

Ritter et al. (2010) stellen in ihrer vergleichenden Promoter- und Enhanceranalyse fest, daß weniger der Homologiewert von zwei Sequenzen, sondern viel mehr ein konserviertes TFBS-Profil zweier Sequenzen, ein Hinweis für eine konservierte genregulative Funktion darstellt. Die Analyse orthologer Enhancerelemente (spezifisch für die Körpergestallt) zwischen Zebrafisch (*Danio rerio*) und Mensch (Trennung der Linien vor ~ 450 Mio. Jahren; Metzger, 2011; Hedges & Kumar, 2003) erbrachte zudem, daß die *cis*-regulative Sequenzevolution der TFBS, einen deutlich größeren Beitrag zu veränderten Genexpressionsmustern leistet, als eine Sequenzevolution der Transkriptionsfaktoren (Ritter et al., 2010). Die Präferenzen der Transkriptionsfaktoren für bestimmte Bindesequenzen haben sich in vielen Fällen seit der Trennung von Invertebraten und Vertebraten kaum verändert. Vermutlich deshalb, da die Veränderung der TF-Bindeaffinität alle Gene, die unter Kontrolle eines TFs stehen, beeinflussen würden (Wray, 2007; Berger et al., 2008; Weirauch & Hughes, 2010).
Mit zunehmender Genomgröße, seit den einzelligen Eukaryoten hin zu komplexen mehrzelligen Organismen, nimmt auch die Anzahl der TFs, z.B. durch Genduplikationen und deren TFBSs zu. Dies birgt wiederum neue Probleme, da die Erkennung einer spezifischen Bindesequenzen unter Tausenden und zum richtigen Zeitpunkt in großen komplexen Genomen für einen einzelnen TF eine Herausforderung darstellt (Pan et al., 2009; Vaquerizas et al., 2009; Stormo & Zhao, 2010; Bird, 2011). Da TFBSs zudem degeneriert und häufig nur 6 bis 8 Nukleotide lang sind, muß also die fehlende TF-Sequenzspezifität über die Kombination mehrerer TFBS zu TF-Modulen wieder hergestellt werden. Eine zweite Möglichkeit ist auch die fehlende Spezifität durch eine gewebespezifische Expression von Proteinkomplexen wieder zu erhöhen (Stormo, 2000; Bird, 2011; Emig & Albrecht, 2011). Dynamische, funktionsspezifische TF-Komplexe, „Enhanceosome“ oder „*cis*-regulative Module“ (CRMs) genannt, können so, über Enhancer- und Promoterbindung, die Genexpression in gewebe- oder entwicklungsspezifischer Weise kontrollieren (Smith et al., 2006; Waleev et al., 2006; Panne, 2008). In einer regulativen DNA-Sequenz können somit mehrere regulative Codes parallel niedergelegt sein, die je nach „Enhanceosom“ in regulative Signale umgesetzt werden.

5.3.1. cTFBS und TF-Module

Die TFBS-Analyse, sowie die Analyse über- und unterrepräsentierter TF-Motive (Tab. 4-11 bis 4-13) und der Vergleich, der in der Genomatix-Datenbank bekannten, funktionellen TF-Module (4-14, 4-15) lassen den Schluß zu, daß die Gene *DDX3X* und *DDX3Y* zu einem gewissen Grad co-reguliert sein könnten. Trotz aller Unterschiede im Aufbau der zwei Promoterregionen scheint ein Grundset an konservierten Faktoren die Allel-spezifischen Sequenzveränderungen überdauert zu haben. Auffällig ist, daß die meisten konservierten TFBSs einen zugehörigen TF mit Hodenexpression haben. Scheinbar bestand schon vor der Trennung der Geschlechtschromosomen ein Zusammenhang zwischen den *DDX3* Promoterregionen und der Keimzellexpression. Eine essentielle Rolle für die Keimzellspezifische *DDX3X/DDX3Y* Expression wird in der konservierten SOX/DMRT TFBS vermutet, die in allen Säugerspezies am 3´Ende der *MSY2/MSY2-X* Sequenz identifiziert werden konnte. Die 19 potentiellen DMRT-TFBS zeigen in der Human *DDX3Y* Promoterregion, im Gegensatz zu *DDX3X*, zudem eine statistisch deutliche Überrepräsentation. Corà et al. (2004, 2005) definierten die Überrepräsentation bestimmter TFBS in Promoterregionen als deutliches Kriterium für eine Funktionalität der beteiligten TFs.

Einen ersten Eindruck eines modularen Aufbaus der Promoterregionen geben die 39 potentiellen *DDX3Y* und 38 *DDX3X* TF-Module (Tab. 4-14 und 4-15). Sechs dieser TF-Module konnten in beiden Promoterregionen identifiziert werden, allerdings meist nicht an homologer Position (Abb. 4-49). Besonders fiel auf, daß viele der ermittelten TFs dieser Analysen eine Expression in Hodengewebe aufweisen. Weiterhin fällt auf, daß sowohl in der *DDX3X* (11x), als auch in der *DDX3Y* (6x) Promoterregion viele potentielle TF-Module mit TFBSs für den Faktor SPI1 (PU.1) kartierbar sind. SPI1 wird in der Keimbahn benötigt, um SSCs und frühe Spermatogonien in undifferenzierten Stadien zu halten (Olive et al., 2007). Ebenfalls in der *DDX3X* Promoterregion sind zehn TF-Module mit SP1 TFBS kartierbar. In der *DDX3X* Promoterregion fallen auch repetitive TF-Module, wie SP1-SPI1 (4x) und SPI1-AP1 (2x) auf, die zwischen *MSY2-X* und *DDX3X* T/B-TSS-II lokalisiert sind. Beide TF-Module kommen in der *DDX3Y* Promoterregion jeweils einmal vor. In der *DDX3Y* Promoterregion sind, durch die *MSY2*-Sequenz Amplifikation, dagegen spezifisch Wiederholungen der SOX-PAX und SOX-OCT TF-Module entstanden.

Die Anwesenheit gemeinsamer TF-Module (Abb. 4-49), wenn auch nicht positionshomolog, unterstützt die Hypothese einer Co-Regulation im selben biologischen Prozeß, der Spermatogenese. Die vermutete Wichtigkeit der X-Y konservierten SOX5, bzw. SPI1 TFBSs wird durch die Beteiligung dieser zwei Faktoren in TF-Modulen unterstrichen (Abb. 4-49). Der Vergleich von TF-Profilen zur Identifizierung gemeinsamer Promotermuster in co-regulierten Genen wird z.B. bei der Suche nach neuen Kandidatengenen für die Alzheimer Krankheit und

weiterer multifaktoriell bedingter Erkrankungen eingesetzt (Park et al., 2002; Döhr et al., 2005; Augustin et al., 2011). Erste Ergebnisse zeigten auch bereits, daß die Transkription der beiden *DDX3* Gene wahrscheinlich nicht ausschließlich in Spermatogonien oder Spermatiden erfolgt, sondern beide Gene in beiden Zelltypen aktiv transkribiert werden (Jaroszynski et al., 2011; Zimmer J., unveröffentlichte Resultate).

Die Beschreibung einer funktionellen HIF1 (Hypoxia inducible Factor 1) TFBS im *DDX3X* Promoter (Botlagunta et al., 2011), bestätigt den in dieser Arbeit angewandten methodischen Ansatz zur Identifizierung potentiell funktioneller TFBS. Das von Botlagunta et al. (2011), in einer MCF 7 Brustkrebszellkultur, experimentell untersuchte „HIF response Element" (HRE), mit der Core-Sequenz „ACGTG" ist 62 bp stromaufwärts von *DDX3X* TSS-I in BAC-Position 30.577 (BAC RP13-13A3) lokalisiert. Die Core-HRE-Sequenz lautet „RCGTG". In der X-Y konservierten Sequenzregion KRiH-8 sind direkt nebeneinander zwei potentielle Core-HREs lokalisiert, die beide auch in der *DDX3Y* Promoterregion konserviert sind (Abb. 4-48). Beide Core-HREs sind in einer längeren TF-Matrix lokalisiert. Diese Sequenzregion scheint also empfänglich für HIF1 Signale. In Tab. 4-10 ist die erste der beiden konservierten Core-HREs „ACACG" in BAC-Position 30.568 (*DDX3Y* BAC-Klon 475I1 = 54.993) angegeben. Das zweite X-Y konservierte Core-HRE (nicht aufgeführt in Tab. 4-10) ist identisch mit dem HRE-1, welches von Botlagunta et al. (2011) untersucht wurde. Das HRE ist ebenfalls in der Primaten Y-Linie (Tab. 4-18) und in der gesamten X-Linie konserviert (Tab. 4-25).
Langanhaltende Hypoxie, z.B. durch Ischämie in Folge einer mechanischen Samenstrangtorsion verursacht, führt im Hodenepithel häufig zu einem irreversiblen Verlust der Keimzellen (Gat et al., 2010). Die Untereinheiten HIF1α, HIF2α und HIF3α bilden unter Hypoxie Heterodimere mit HIF1β (= Aryl Hydrocarbon Receptor Transporter, ARNT) und binden an das HRE von Hypoxie-regulierten Genen (Botlagunta et al., 2011). Im Hodengewebe kommt zusätzlich zu den ubiquitär exprimierten Untereinheiten eine hodenspezifische HIF1α-Testis Untereinheit mit dominant-negativer Funktion vor (Lysiak et al., 2009). Funktionelle Bedeutung für den HIF1-Komplex bei der Antwort auf einen veränderten Sauerstoffgehalt in Hodengewebe, konnte in Mensch, Maus und Ratte gezeigt werden (Lysiak et al., 2009; Paul et al., 2009; Gat et al., 2010; Palladino et al., 2011). Eine mögliche Beteiligung und erhöhte Expression von DDX3X in einem Streß-induzierten Signalpathway, wie dem Sauerstoffmangel, ist interessant, da die DDX3 Proteine als zusätzliche Helikase bei der Translationsinitiation von mRNAs mit langen und strukturierten 5´UTRs, wie z.B. dem Zellzyklus-Regulator Cyclin-E1, Verwendung finden (siehe Kapitel 1.2.). Die DDX3 Proteine könnten so zu einer Stabilisierung der Translation und des Zellzykluses (G1-Phase), in Situationen von Streß-induzierter Mangelerscheinung im Hodenepithel, führen. Genauso könnte aber auch ein Zellzyklusarrest die Folge einer erhöhten DDX3 Expression sein, da DDX3X an die Promotersequenz von p21 binden kann (Chao et al.,

2006; Kapitel 1.2.). Die durch p21 induzierte Streß-Antwort inaktiviert wiederum CDK2-Komplexe und führt so zu einem Arrest in der G1-Phase (Chao et al., 2006). Welcher dieser zwei Signalwege in Keimzellen, in Folge einer erhöhten DDX3 Expression, genutzt werden kann, ist bisher allerdings unbekannt.

In Kapitel 4.6.3.2. fiel besonders die statistische Überrepräsentation von E2F-Faktoren in beiden *DDX3*(*X/Y*) Promoterregionen auf (Tab. 4-11 bis 4-13). Spezifisch für *DDX3X* konnten zudem potentielle TF-Module mit E2F-Faktoren identifiziert werden (Tab. 4-15). E2F Transkriptionsfaktoren kontrollieren essentiell die zelluläre Proliferation und werden in aktivierende und reprimierende E2Fs eingeteilt (Pohlers et al., 2005; Berckmans & De Veylder, 2009; Chen et al., 2009; Aguilar & Fajas). In der Maus wird z.B. E2F6 Keimzell-spezifisch exprimiert (Pohlers et al., 2005).
In der vertikalen *DDX3X* X-Linie wird eine E2F TFBS von Maus bis Mensch konserviert gefunden. Diese überlappt mit dem *DDX3X* T/B-TSS-II DPE-Motiv (Tab. 4-25, Abb. 4-52, A-4). Stromaufwärts zu *DDX3Y* TSS-I-(T) sind in der Primaten Y-Linie und im Rind im Intron-T ebenfalls konservierte E2F TFBS vorhanden. Auf Grund der E2F TFBS-Lokalisation in den proximalen *DDX3*(*X/Y*) Promoterabschnitten besteht die Möglichkeit, daß die Transkripte mit kurzer 5´UTR einer Zellzykluskontrolle unterstellt sind. Ob die (C,G)-reichen E2F TFBSs, wie auch die SP1 TFBSs, für die *DDX3X* und *DDX3Y* Genexpression funktionell sind, oder die beobachtete TFBS-Überrepräsentation eine Folge der CpG-Islands ist, konnte bisher experimentell nicht determiniert werden. Die Funktion der DDX3 Proteine im Zellzyklus sind in Kapitel 1.2. beschrieben.

Einen weiteren Hinweis für eine konservierte Zellzykluskontrolle der *DDX3*(*X/Y*) Gene könnte das CHR TF-Bindemotiv (“cell cycle genes homology region”) in der X-Y konservierten Sequenz KRiH-2 (Tab. 4-10, Abb. 4-48) geben. CHR-Motive werden häufig in Promotersequenzen von Genen mit hoher Expression in der S- und G_2-Phase des Zellzykluses gefunden (Müller & Engeland, 2009). In der vertikalen *DDX3X* Linie, aber nicht in der Y-Linie, konnte das CHR-Motiv zudem von Maus bis Mensch konserviert identifiziert werden (Tab. 4-25). Gene, die unter Kontrolle eines CHR-Motives stehen, werden meist in der G_0- und G_1-Phase reprimiert. D.h., daß die CHR-regulierten Gene während des Zellzyklus in Wellen exprimiert werden. Welche Faktoren an das CHR-Motiv binden können, ist bisher nicht genau bekannt (Müller & Engeland, 2009). Die Expression CHR-regulierter Gene wird durch CCAAT-Boxen aktiviert, an die der TF „Nuclear factor-Y“ (NF-Y) bindet, der auch die Histonacetylierung und die Rekrutierung der Histonvariante H2A.Z, im Bereich der TSS, beeinflussen kann (Hou et al., 2010; Gatta & Mantovani, 2011). Potentielle CCAAT-Boxen wurden nur in der *DDX3Y* Promoterregion in den Positionen 53.174 (stromaufwärts zu KRiH-2) und 53.333 (innerhalb KRiH-2) identifiziert.

Reverse CCAAT-Boxen „ATTGG“, die häufig in TATA-losen Promoterregionen vorkommen (Mantovani, 1998), konnten weder im *DDX3Y*, noch im *DDX3X* Promoter kartiert werden. Da die zwei CCAAT-Boxen allerdings weit außerhalb der *DDX3Y* Core-Promoterregionen liegen, ist es fraglich, ob sie tatsächlich als funktionelle Promoterelemente agieren. Meist kommen diese CCAAT-Boxen im proximalen Promoter im Abstand von -60 bis -100 zu einer TSS vor (Mantovani, 1998; Dolfini et al., 2009; Martinez, 2010). Der potentielle Einfluß von reprimierenden CHR-Motiven und aktivierenden CCAAT-Boxen auf die *DDX3*(*X/Y*) Expression ist experimentell bisher nicht untersucht. Von den DDX3Y und DDX3X Proteinen ist bisher bekannt daß sie vor allem in der G_1-S Phase des Zellzykluses aktiv sind (Fukumura et al., 2003; Lai et al., 2010).

5.3.2. CpG-Islands

Die Anwesenheit eines CpG-Islands (CGI) in beiden Promoterregionen und der konservierte Abstand vom Translationsstartcodon ATG bis zur konservierten HIF1 TFBS unterstützen den Eindruck (Abb. 4-49), daß die proximalen *DDX3*(*X/Y*) Promoterregionen generell konservierter sind, als die distalen Keimzell-spezifischen Promoterregionen. Das distale *DDX3X* CpG-Island (CpG-P1) fehlt gänzlich in der *DDX3Y* Promoterregion (Abb. 4-47). Der (C,G)-Gehalt der *DDX3Y* Promoterregion ist reduziert, da der Vergleich von Human *DDX3X* zu Opossum *Ddx3* eindeutig gezeigt hat, daß im Ursprungszustand eines potentiellen Säuger Ur-*Ddx3* Gens ein langer (C,G)-reicher Sequenzabschnitt in der Promoterregion vorlag (Abb. 4-46).

Die Reduzierung des (C,G)-Gehaltes in der Human *DDX3Y* Promoterregion scheint durch die Vervierfachung des *MSY2*-Repeats begünstigt worden sein. Die 409 Nukleotide des Human *MSY2* Repeats (BAC-Klon 475I1, Pos. Pos. 53.854 - 54.262) haben zusammen einen geringen (C,G)-Gehalt von nur 45,23%, im Vergleich zu den 65,42% der *MSY2-X* Sequenz (BAC-Klon RP13-13A3, Pos. 29.570 - 29.676). Die Veränderung der Nukleotidzusammensetzung zu der insgesamt (A,T)-reichen *MSY2*-Minisatellitensequenz, könnte wiederum die Entstehung der multiplen TATA-Motive innerhalb des *MSY2* Repeats begünstigt haben.

In der untersuchten 2,5 kb *DDX3X* Promoterregion erscheint die *MSY2-X* Sequenz gleich einer Abgrenzung zwischen der (A,T)-reichen distalen Promoterregion stromaufwärts von *MSY2-X* und der (C,G)-reichen Region stromabwärts von *MSY2-X*. Da bekannt ist, daß (C,G)-reiche Promoterdomänen, über spezifische CpG-Methylierungsmuster, aktive Promoterelemente bilden können, besteht für die *DDX3X* Promoterregion die Möglichkeit einer umfassenden Methylierung/Demethylierung der Promoterregion. Zudem ist bekannt, daß die Chromatinstruktur in CpG-Islands generell weniger stabil ist und daher für die Transkriptions-maschinerie leichter zugänglich ist (Deaton & Bird, 2011). Dies hätte wiederum großen Einfluß auf die Chromatinstruktur und Transkription von *DDX3X*, die damit deutlich unterschiedlich zu *DDX3Y* erscheint. In Kapitel 4.2 konnte gezeigt werden, daß in der Region zwischen *DDX3X* T-

TSS-III und TSS-I mehrere TSS von DBTSS angegeben werden, die keinem Core-Promotermotiv zuzuordnen waren (Tab. 4-4). Das legt die Vermutung nahe, daß, wegen des ausgedehnten CpG-Islands in der *DDX3X* Promoterregion, alternative Transkriptionsinitiations-mechanismen z.B. über den TF SP1, mit Bindung an mehrere GC/GT-Boxen (BAC-Klon RP13-13A3, Pos. 29.681 - 29.932), entstanden sein könnten (siehe Kapitel 4.6.3).
SP1 Bindestellen gelten zudem als Methyl-sensitiv und ermöglichen somit eine gewebespezifische Genexpression. Für *DDX3X* konnten fünf SP1 TFBS (BAC-Klon RP13-13A3, Pos. 29.681 - 29.932 = 252 bp) in den (C,G)- und Pyrimidin-reichen Bereich stromabwärts von *MSY2-X* kartiert werden (Abb. 4-48, 4-58, A-17). Der Faktor SP1 scheint zudem in mehreren der kartierten TF-Module in der *DDX3X* Promoterregion beteiligt zu sein (Tab. 4-15). Durch gewebespezifische Methylierung der CpG-Dinukleotide in den SP1-TFBSs, könnte die Keimzell-spezifische *DDX3X* T-TSS-III Promoterregion reguliert werden und die SP1-Bindung an die Promoterregion könnte vor einer *de novo* Methylierung schützen (Deaton & Bird, 2011). Neben einer Methylierung/Demethylierung eines kompletten CGIs scheint auch die Methylierung einzelner wichtiger Methyl-sensitiver TFBS in kritischen Promoterbereichen für eine Veränderung der Genexpression auszureichen (Cao et al., 2000; Pogribny et al., 2000).

Weitere Methyl-sensitive TFBS mit CpG-Dinukleotid sind z.B. die „half-CRE-Sequenz“ (CGTCA; cAMP response element) und das „half-C/EBP-Element“ (CGCAA; CCAAT/enhancer-binding protein), welche von unterschiedlichen Faktoren je nach Methylierungszustand erkannt werden (Rishi et al., 2010). Die unmethylierte CRE-Sequenz wird von CREB, die methylierte von C/EBPα erkannt (Rishi et al., 2010). Das methylierte „half-C/EBP-Element“ wird ebenfalls von C/EBPα erkannt. In der *DDX3X* Promotersequenz sind zwei „half-C/EBP-Elemente“ in Position 29.521 (5´ zu *MSY2-X*) und 30.334 (3´ zu T/B-TSS-II) lokalisiert. Dagegen sind in der *DDX3Y* Promotersequenz drei „half-CRE-Sequenzen“ innerhalb des *MSY2* Repeats in den Positionen 54.028, 54.126 und 54.225 lokalisiert. Diese zwei Sequenzelemente, die jeweils spezifisch in der *DDX3X* oder *DDX3Y* Promoterregion vorkommen, sollten also die Möglichkeit einer gewebespezifischen Aktivierung/Repression je nach Methylierung der CpG-Dinukleotide bieten.
Insgesamt entsteht also ein Bild, daß neben den konservierten TFs und TF-Modulen, die z.B. die Expression in Keimzellen und eine Abhängigkeit der Zellzykluskontrolle regulieren könnten, die unterschiedlichen Core-Promotermodule, unterschiedlichen Sequenzeigenschaften wie CpG-Islands und Allel-spezifischen TF-Bindestellen zu den divergenten *DDX3X* und *DDX3Y* Expressionsmustern im Menschen führten. Im Gegensatz zum eher konservativ evolvierenden *DDX3X* Promoter, auf Grund der regulären X-Chromosomrekombination im weiblichen Geschlecht, wurde der *DDX3Y* Promoter in der Säugerevolution stark verändert und hat vor allem in der Primatenlinie männlich spezifische Eigenschaften entwickelt.

5.4. Potentiell funktionelle Unterschiede Chromatinmodule für die hodenspezifische Transkriptionsaktivierung der *DDX3X* und *DDX3Y* Genloci

5.4.1. *DDX3X* Genexpressionskontrolle durch ZF5 Homomer-Chromatinkomplexe

Auf einen deutlichen Unterschied der Expressionskontrolle von *DDX3X* und *DDX3Y* weist auch die *DDX3X*-spezifische POZ-Domäne (BAC-Klon RP13-13A3, Pos. 29.552 - 29.812 = 261 bp) hin. In der Analyse der überrepräsentierten Sequenzmotive fielen 12 potentielle ZF5 Bindestellen durch einen besonders hohen Z-Score auf (Tab. 4-12, Abb. 4-58, Tab. A-17). Auf Grund der Lage der repetitiven ZF5 TFBSs, in dem S-Trakt zwischen *MSY2-X* und *DDX3X* T-TSS-III, könnte eine Öffnung und Schließung der hodenspezifischen *DDX3X* Promoterdomäne, auch durch eine funktionelle Chromatinstruktur mit ZF5-Homomerkomplexen, diskutiert werden.

Ein gut untersuchter Vertreter solch einer BTB/POZ TF-Funktion ist der *Drosophila* „GAGA-binde Faktor“ (GAF), als ein Bestandteil des *Drosophila* Trithorax-Komplexes. GAF bindet als GAF-Homomerkomplex an wiederholt hintereinander liegende „GAGA“-Sequenzmotive. Dadurch kann lokal eine Chromatinstruktur ohne Histone etabliert werden, indem ein Histonoktamer verdrängt und eine alternative Promotertopologie über den GAF-Komplex etabliert wird (Abb. 5-1; Tsukiyama et al., 1994; Katsani et al., 1999; Pessler & Hernandez, 2003; Lehmann, 2004; Adkins et al., 2006; Kypr et al., 2007; Matharu et al., 2010).

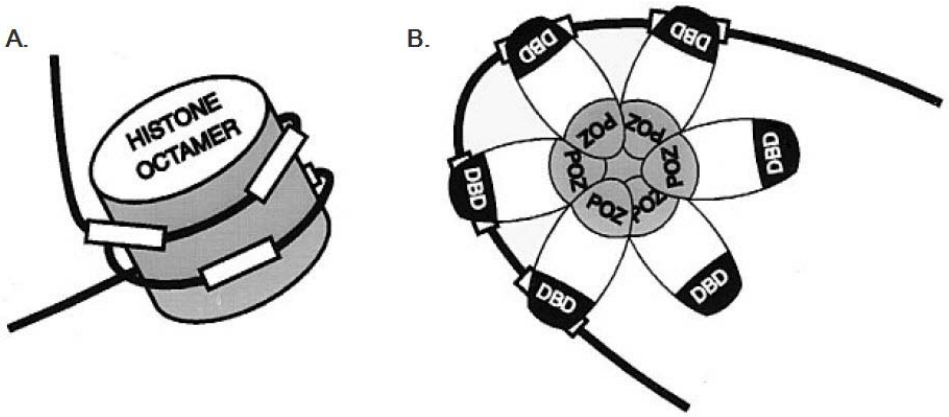


Abb. 5-1: Mögliche alternative Chromatinstruktur nach lokaler Verdrängung eines Nukleosoms durch einen GAF-Komplex.
A: Nukleosomenmodell mit Histon-Oktamer. Leere Boxen stellen GAGA-Motive dar. **B:** Die Multimerisierung der GAFs erfolgt über die POZ-Domänen, die DNA-Bindung erfolgt über die Zinkfinger-Domänen. Das Histonoktamer wird durch den GAF-Komplex lokal verdrängt. Lokal bildet sich hierdurch eine alternative Chromatinstruktur aus (nach Katsani et al., 1999).

Als Vertebraten-Homolog zu *Drosophila* GAF konnte das *c-Krox* Gen (= ZFP67) identifiziert werden (Matharu et al., 2010). Funktionelle „GAGAG-Boxen“ in Bezug auf die Chromatinstruktur konnten z. B. in den Vertebraten *HOX*-Gen Klustern gezeigt werden (Matharu et al., 2010).

Das Human ZF5-TF Protein (= ZFP161) hat, wie der *Drosophila* GAF, ebenfalls eine N-terminale BTB/POZ-Domäne und C-terminale Zinkfinger-Domänen (Sugiura et al., 1997). ZF5 geht sowohl als Homomer, als auch als Heteromer Proteininteraktionen ein (Numoto et al., 1999). Die ZF5 Konsensus-Bindesequenz (5´-GC/GGCGCGA/G-3´) wurde von Obata et al. (1999) festgestellt. Im Unterschied zu GAF bindet ZF5 also nicht Purin-reiche Sequenzen, sondern (C,G)-reiche DNA-Abschnitte. In der Maus ist eine X-chromosomale *Zfp161-rs1* Genkopie mit Keimbahn-spezifischer Funktion bekannt (Sobek-Klocke et al., 1997; Mueller et al., 2008; Stouffs et al., 2009). Im Maus *Tescalin* Genpromoter (*Tesc*), welches in männlichen Keimzellen während der Keimzelldifferenzierung exprimiert wird, scheint ein Zf5-Faktor kritisch für die Aktivierung von *Tesc*, zusammen mit SP1-Bindemotiven, zu sein (Perera et al., 2001, 2010). Es ist somit durchaus möglich, daß ZF5 auch in der Human Spermatogenese eine wichtige Funktion für die Keimzell-spezifische Aktivierung von Genen, wie *DDX3X*, hat.

5.4.2. *DDX3Y* Genexpressionskontrolle durch S/MARs

Die Analyse potentieller S/MAR-Elemente in Kapitel 4.8.2. zeigte deutliche Unterschiede in der Anzahl zwischen den *DDX3Y* und *DDX3X* Genloci. In der untersuchten 50 kb langen Y-chromosomalen Region (GenBank Acc.No. NC_000024.9, Position 14.998.955 - 15.048.954) konnten, mit den zwei verwendeten Programmen MARFinder (3x) und SMARTest (8x), insgesamt 10 potentielle S/MAR-Elemente kartiert werden. Allerdings konnte nur ein potentielles S/MAR-Element (Y-S/MAR Nr. 5 u. Y-S/MAR-42.700), am 3´ Ende des *CASKP1* Gens, zwischen den *DDX3Y* und *CASKP1* Transkriptionseinheiten, von beiden Programmen parallel vorhergesagt werden (Tab. 4-26, Abb. 4-55). Für die untersuchte 50 kb lange X-chromosomale Region (GenBank Acc.No. NC_000023.10, Position 41.175.840 - 41.225.839) konnten insgesamt nur drei potentielle S/MAR-Elemente kartiert werden. Es fällt also auf, daß in der Y-chromosomalen Region wesentlich mehr Sequenzen vorkommen, die *in-silico* den verschiedenen S/MAR-Eigenschaften entsprechen. Die experimentelle Bestätigung einer funktionellen intronischen S/MAR-Sequenz im Human „Phosphoinositide 3-Kinase γ“ Gen (*PI3Kγ*) zeigt, daß eine solche kombinierte *in-silico* Kartierung (SMARTest + MARFinder) tatsächlich zur Identifizierung von S/MAR-Elementen führen kann (Dai et al., 2006; Cai et al., 2010).

Schübeler et al. (1996) beschreiben in stabilen Transfektionsexperimenten einen Entfernungs-effekt der S/MARs in Abhängigkeit des Promoterkontextes. In einem Abstand von 4 kb und mehr zu einer TSS konnte eine aktivierende Wirkung auf die RNA-Expression festgestellt werden. Im Abstand von weniger als 2,5 kb wurde eine fast vollständige Unterdrückung der Expression festgestellt. Das ~ 4,8 kb stromaufwärts des *DDX3Y* ATGs lokalisierte Y-S/MAR-Element Nr. 2 könnte somit einen aktivierenden Effekt auf die *DDX3Y* Genexpression haben.

Die gewebespezifische Expression der Human Tyrosin-Hydroxylase (TH) z.B. steht im Zusammenhang mit S/MARs im TH-Genlocus, die eine differentielle Expression regulieren (Lenartowski & Goc, 2002). Boulikas (1995) gibt eine Übersicht zu gewebespezifischen S/MARs und deren Einfluß auf die Genaktivität benachbarter Gene. Ganapathi et al. (2005) haben in Human eine genomweite Suche nach potentiellen S/MARs durchgeführt und haben eine Häufung von S/MARs in gewebespezifisch exprimierten Genloci, im Vergleich zu Housekeeping-Genen, identifiziert. Die unterschiedliche Anzahl an S/MARs in den *DDX3X* und *DDX3Y* Loci könnte somit zu einer unterschiedlichen Expression der Gene in den verschiedenen Humangeweben führen.

Die Replikation der DNA erfolgt in höheren Eukaryoten in Assoziation mit der Kernmatrix (DePamphilis, 1999). Konsequenterweise sind „Origin of Replications“ (ORIs) daher auch mit S/MARs ko-lokalisiert. Experimentell sind S/MARs in klassischen Reportergenvektoren, im Unterschied zu Enhancer-Elementen, nur in stabilen Transfektionsexperimenten analysierbar (Schübeler et al., 1996). Alternativ dazu wurden in den vergangenen Jahren verschiedene „selbstreplizierende episomale Vektoren“ (REV), auch „non-viral episomal plasmid DNA (pDNA)“ genannt, entwickelt (Tessadori et al., 2010; Wong et al., 2011). Bei Anwesenheit eines funktionellen S/MAR-Elementes in einer solchen extra-chromosomalen Einheit werden virale Sequenzen im Vektor (SV40, EBV), zur DNA-Replikation und Aufrechterhaltung der exogenen Genexpression, überflüssig (Wong et al., 2011). Da die so gestalteten Episome nicht ins Genom der Wirtszelle integrieren und somit ein geringes onkogenes Potential besitzen, können schädliche Mutationen vermieden werden. Inzwischen sind verschiedene „S/MAR Minicircle“ Vektorsysteme erhältlich, die auf einem solchen Episom basieren. Die Aufrechterhaltung der Genexpression in Zellkultur wird mit Luziferase oder eGFP (enhanced green fluorescent protein) kontrolliert.

Für S/MAR-Elemente in Episomen, z.B das S/MAR-Element des β-Interferon-Genlocus (IFN-SAR), konnte in Zellkulturexperimenten gezeigt werden, daß sie *de novo* Methylierungen der dazugehörigen Vektor-Promoterregionen verhindern können (Dang et al., 2000; Gill et al., 2009). Die exogene Genexpression kann somit durch ein S/MAR-Element aufrechterhalten werden. Falls eine solche „Anti-Methylierungsfunktion“ von S/MAR-Elementen auch nativ, und nicht nur in einem Episom, existiert, könnte sie z.B. die Keimzell-spezifische *DDX3Y* Promoterdomäne in Keimzellen in einem aktiven, geöffneten Zustand halten. Da die *DDX3Y* Promoterregion, auf Grund des fehlenden langen CGIs, vermutlich keiner kompletten Methylierung/Demethylierung unterliegt, könnte das stromaufwärts des *DDX3Y* ATGs lokalisierte Y-S/MAR-Element Nr. 2, aber die Methylierung einzelner wichtiger Methyl-sensitiver TFBS, in den kritischen Promoterbereichen, verhindern. Die kartierten S/MAR-Elemente in den *DDX3Y* (Abb. 4-55) und *DDX3X* (Abb. 4-56) Genloci könnten also in einem episomalen

Vektorsystemen auf Funktionalität getestet werden. Allerdings sollte die Analyse in mehreren Zellinien erfolgen, da bekannt ist, daß S/MARs nicht in allen Zellinien gleichermaßen funktionieren (Papapetrou et al., 2006).

5.4.3. S/MARs und deren Assoziation zur Kernmatrix

Eine weitere Möglichkeit zur differentiellen, gewebespezifischen Genkontrolle über aktive und inaktive Chromatinregionen bietet die Assoziation des Chromatins mit strukturgebenden Proteinen der Kernmatrix, den Intermediärfilament-bildenden Laminen A, B, C (Abb. 5-2; Schütz et al., 2005; Barendziak, 2006; Heessen & Fornerod, 2007; Coutinho et al., 2009; Prokocimer et al., 2009).

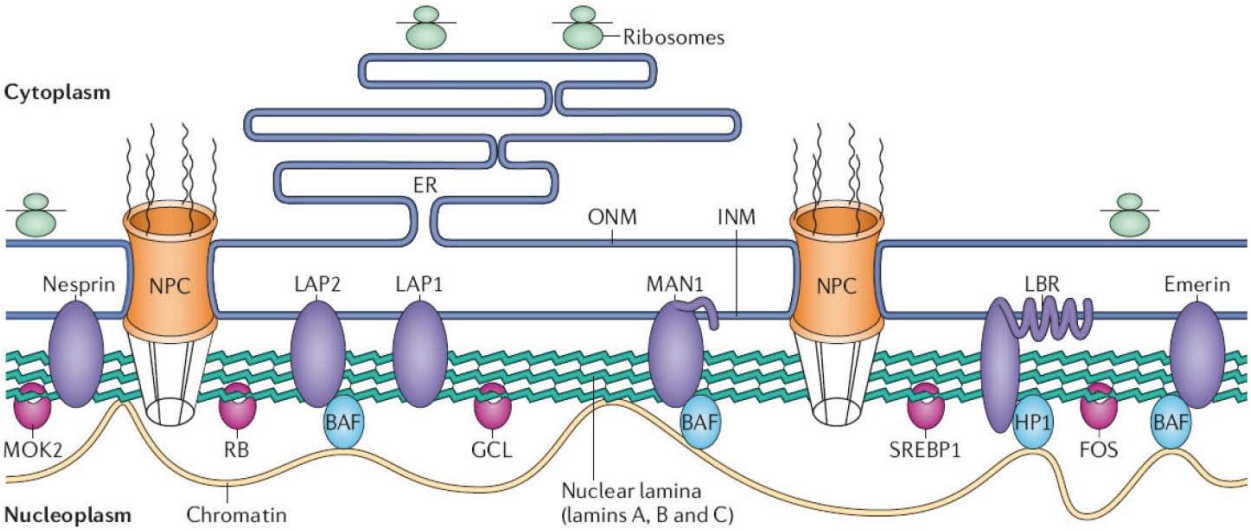


Abb. 5-2: Struktur und Funktion der Kernlamina.

Die Kernlamina (= grün) schließt an die innere Kernhülle („inner nuclear membrane“ = INM) an und hilft die Kernstabilität aufrecht zu erhalten, das Chromatin zu organisieren und die Kernporen („nuclear pore complexes“ = NPCs) zu binden. Zudem dient die Kernlamina einer Vielzahl von Proteinen („nuclear envelope proteins“ = lila) und TFs (= pink) als Anker. Einige dieser Hüllproteine sind: Nesprin, Emerin, „Lamina-assoziierte Proteine 1 und 2“ (LAP1, LAP2), der „Lamin B Rezeptor“ (LBR) und das „Inner nuclear membrane protein MAN1“ (= *LEMD3*). TFs, die an die Lamina binden sind z.B.: das „Retinoblastoma transcriptional regulator“ (RB), das BTB/POZ-Protein „Germ cell-less Protein-like 1“ (*GCL*), das „Sterol response element binding Protein“ (SREBP1), das Proto-Onkogen c-FOS (*FOS*) und das Zinkfinger Protein ZNF239 (= *MOK2*). Der Faktor „Barrier to autointegration factor“ (BAF) ist ein Chromatin-assoziiertes Protein, welches sowohl an die Kernlamina, als auch an einige der zuvor genannten Hüllproteine binden kann. „Heterochromatin Protein 1“ (HP1) bindet „Chromatin Lamin B Rezeptor“ (LBR). ONM = „outer nuclear membrane“ (nach Capell & Collins, 2006).

Die funktionelle Bedeutung der Kernlamina zeigt sich in einer Gruppe von Pathologien, den Laminopathien, die z.B. durch Mutationen im Gen *LMNA* (Lamin A) ausgelöst werden und so zu einer fehlerhaften Kern- und Chromatinorganisation führen (Coutinho et al., 2009). Auch konnte gezeigt werden, daß S/MARs von Laminen gebunden werden und so zur Etablierung aktiver und inaktiver Genloci beitragen (Ludérus et al., 1994). Dabei können gewebespezifisch Domänen etabliert werden, da der Lamin-Typ B3, eine Splicevariante des Typs B2, spezifisch in

postmeiotischen Keimzellen der Säugetier exprimiert wird, und so eine spezifische Reorganisation des Chromatingerüstes in Keimzellen ermöglicht wird (Schütz et al., 2005; Barendziak, 2006; Shevelyov et al., 2009; von Moeller et al., 2010).
Der Lamina-assoziierte TF „Germ cell-less Protein-like 1“ (*GCL*) stellt eine Verbindung zwischen Kernmatrix, Chromatinkondensation, Zellzykluskontrolle und Keimzellen her (Margalit et al., 2005). GCL ist in bestimmten fleckenförmingen „speckles“, membranlose Reaktionsräume von Proteinkomplexen im Zellkern, an der inneren Kernhülle lokalisiert (de la Luna et al., 1999). Eine Keimzellfunktion für den BTB/POZ-Faktor GCL, auch DIP (DP-interacting Protein) genannt, konnte in *Caenorhabditis elegans*, *Drosophila* und Maus gezeigt werden (Kimura et al., 1999; Leatherman et al., 2000; Holaska et al., 2003). Die negative Zellzykluskontrolle von GCL erfolgt über die Interaktion mit E2F-Proteinkomplexen, indem diese durch GCL an die innere Kernhülle rekrutiert und inaktiviert werden (de la Luna et al., 1999).

Das gehäufte Vorkommen potentieller S/MAR-Elemente im *DDX3Y* Genlocus könnte somit auch Bedeutung für die Kondensation des Chromatins im Human AZFa-Locus haben. In der Meiose wird das Chromatin der gepaarten Gonosomen kondensiert und im MSCI inaktiviert (Turner, 2007; siehe Kapitel 1.4.). Für die AZFb- und AZFc-Loci, die wie AZFa in der chromosomalen Region Yq11 lokalisiert sind, wird, neben der Funktion der dort kodierten Proteine, auch ein Chromatin-Faltungscode diskutiert, der für die Y-Chromosom Kondensation in der Keimbahn notwendig ist (Vogt et al., 2008). Bekannt ist auch, daß das Y-Chromatin in unterschiedlichen Zelltypen und Lebensabschnitten verschiedene Kondensationszustände aufweisen kann. Das Human Y-Chromosom weist vor der Pubertät einen dekondensierten Zustand in Yq11 in prämeiotischen Keimzellen auf, der nach der Pubertät dann in Sertoli Zellen zu sehen ist (Speed et al., 1993; Armstrong et al., 1994).

5.4.4. S/MARs und *DDX3*(*X/Y*) 3´UTR-Bildung

Die Lage des potentiellen Y-S/MAR-Elementes Nr.4 (Tab. 4-26) zwischen den proximalen *DDX3Y* PAS1 und PAS2 und der distalen PAS3 könnte Implikationen für die gewebespezifische Auswahl der Polyadenylierungsstellen haben. Gleiches gilt für das homologe X-S/MAR-Element Nr. 1 (rel. Pos. 32.656) im *DDX3X* Exon17 (Abb. 4-56).
Von MARBPs wie SAFB1 ist bekannt, daß sie S/MARs, Chromatinarchitektur, Transkriptionsmaschinerie und Splicingkomplexe räumlich zusammenbringen können (Renz & Fackelmayer, 1996; Nayler et al., 1998; Cai et al., 2003; Li et al., 2003; Galande et al., 2007; Ottaviani et al., 2008; Garee & Oesterreich, 2010). Es besteht also die Möglichkeit, daß durch spezifische Histonmodifikationen und Veränderung der Chromatinarchitektur, die Bildung alternativer 3´ Transkriptendbildungen kontrolliert werden kann (Lian et al., 2008; Lutz, 2008; Spies et al., 2009; Schwartz & Ast, 2010). MARBPs wie SATB1, SATB2 und SMAR1 werden nicht ubiquitär

exprimiert (HPA; Wang et al, 2010) und könnten so alternative Polyadenylierungen gewebespezifisch steuern. Für SATB2 z.B. konnte, per Antikörperfärbung, nur in 9 von 65 analysierten Zelltypen (u.a. Nervenzellen im Cortex, Keimzellen im Hodenepithel) eine Expression dargestellt werden (Human Protein Atlas, HPA).

5.4.5. *Alu*-Sequenzen und *DDX3X* 3´UTR-Bildung

Die Integration von Retrogenen und generell von transposablen DNA-Sequenzen in das X-Chromosom kam während der Säugerevolution, im Vergleich zu den Autosomen, überproportional häufig vor (Emerson et al., 2004). Es stellt sich somit die Frage, ob die Ansammlung multipler *Alu*-Elemente, zwischen den *DDX3X* Exon17a, 17b und 17c Positionen, Einfluß auf das Transkriptionsmuster von *DDX3X* haben kann (Abb. 4-53)? Einen Hinweis auf diese Hypothese liefert die hodenspezifische Expression der *DDX3X* Transkripte mit Exon17a/b/c 3´UTR-Splicevariante (Kapitel 4.8.1.).

So konnte gezeigt werden, daß der Methylierungszustand von *Alu*-Elementen in der männlichen Keimbahn und im Soma voneinander abweicht (Chesnokov & Schmid, 1995). Unterschiedliche *Alu*-Methylierungsmuster in Soma und in Keimzellen und damit verbunden kondensierte und dekondensierte Chromatinkonformationen (Rubin et al., 1994; Brohede & Rand, 2006; Kim et al., 2007), könnten somit zu einer hodenspezifischen Nutzung der distalen *DDX3X* Polyadenylierungsstelle PAS5 führen. Die Transkription wäre folglich nur in einer unmethylierten und dekondensierten Chromatinkonformation möglich. Wood et al. (2008) haben ähnliche Prozesse für Imprintingregionen mit differentieller Methylierung zwischen paternalem und maternalem Allel gezeigt. Ein möglicher Faktor, der *in-vitro* für eine Hypomethylierung von CpG-Dinukleotiden in *Alu*-Elementen ausreichend ist, wurde als „Sperm *Alu*-binding Protein“ (SABP) beschrieben (Chesnokov & Schmid, 1995). SABP bindet selektiv an *Alu*-Elemente da eine generelle Hypomethylierung *in-vitro* nicht festgestellt werden konnte (Chesnokov & Schmid, 1995). In Abwesenheit von SABP könnten CpG-Dinukleotide in *Alu*-Elementen durch nukleäre Proteine mit „Methyl-CpG binding Domain“ (MDB), wie MeCP2, erkannt werden. MeCP2 rekrutiert durch die Bindung an mCpG-Dinukleotide weitere Proteinkomplexe, wie den repressiven Histon-Deacetylasekomplex, und vermittelt so eine transkriptionelle Repression der entsprechenden genomischen Regionen (Bird, 2011).

Weiterhin deutet der RT-PCR-Vergleich, von *DDX3X* Exon17 Transkriptvarianten, mit dem Schimpansen daraufhin (Abb. 4-54), daß in dieser *Alu*-reichen Region auch speziesspezifisch neue Exone entstehen können. Die Rekrutierung von *Alu*-Elementen ist ein bekannter Mechanismus, wie neue Sequenzen in Transkriptionseinheiten, meist durch alternatives Splicing, integriert werden können (Cordaux & Batzer, 2009). Die Anwesenheit von *Alu*-Elementen in Intronsequenzen kann allerdings auch zu Elongationsfehlern, auf Grund instabiler

Primärtranskripte, führen, die einen vorzeitigen Abbruch der Transkription zur Folge haben können (Cordaux & Batzer, 2009).

Es stellt sich zudem die Frage nach dem Zeitfaktor bei einer Transkription über ~ 31.100 bp (*DDX3X* T-TSS-III bis PAS5)? Das zweite hodenspezifische *DDX3X* Transkript, mit ~ 14.700 bp (*DDX3X* T-TSS-III bis PAS1) benötigt, bei einer Elongationsleistung in Eukaryoten von ~ 20 Nukleotiden pro Sekunde (Castillo-Davis et al., 2002), weniger als die Hälfte der Elongationszeit von knapp 26 min für das lange *DDX3X* Primärtranskript mit Exon17a/b/c. Der hoden- und ovarspezifische Transkriptionselongations Faktor S-II (*TCEA2*) könnte dazu beitragen (Umehara et al., 1995), die langen *DDX3X* Primärtranskripte in Keimzellen zu stabilisieren, um *Alu*-bedingte Elongationsfehler zu reduzieren.

Zusammenfassend können für die *DDX3Y* und *DDX3X* Genexpression zwei unterschiedliche Modelle der Expressionskontrolle vorgeschlagen werden. Für die *DDX3Y* Genkontrolle scheint eine gewebespezifische Etablierung von Chromatinschleifen über S/MARs möglich zu sein. Im Unterschied dazu, könnte als funktionelle Basisstruktur für die *DDX3X* Expressionskontrolle eine gewebespezifische Methylierung und Demethylierung von CpG-Islands, sowie *Alu*-Elementen, mit lokal begrenzten Chromatinstrukturen über ZF5 Homomerkomplexe, etabliert worden sein. Unterstützt wird diese Hypothese einerseits durch Anzahl und Lage der (C,G)-reichen Sequenzen in der *DDX3X* Promoterregion, sowie methylierbaren *Alu*-Sequenzen im gesamten *DDX3X* Genlocus und andereseits durch die Anzahl an S/MAR-Elementen im *DDX3Y* Genlocus (Abb. 4-53).

5.5. Funktionelle *DDX3X* Promoteranalysen

Generell sind Enhancerelemente in der Lage die Transkription eines Reportergenkonstruktes zu regulieren, egal in welcher Position, Entfernung, oder Orientierung zum experimentellen Promoter das Element positioniert wird (Kleinjan & van Heyningen, 2005; Bulger & Groudine, 2010). Die meisten Enhancerelemente können in Reportergenkonstrukten die Transkription aktivieren, auch wenn Promoter und Enhancer *in vivo* nicht zum selben Genlocus gehören und nicht dasselbe Expressionsmuster aufweisen (Kain & Ganguly, 2001; Jiang et al., 2008; Bulger & Groudine, 2010).

5.5.1. Kartierung *cis*-regulativer Module in der *DDX3X* Promoterregion

Für die 2,5 kb *DDX3X* Promoterregion wurde anhand von zehn pGL4 Promoter-Reportergenkonstrukten, mit sequentiell verlängerten Promotersequenzen, ein Aktivitätsprofil

erstellt (Abb. 4-59). Die höchste Luziferaseaktivität wurde in Konstrukt Nr. VI festgestellt, welches die 5´ Region von *DDX3X* Exon1 bis einschließlich *MSY2-X* beinhaltet (Abb. 4-59). Zwei weitere Peaks wurden in Konstrukt Nr. II und Nr. VIII gemessen. Das wellenförmige Aktivitätsprofil deutet daraufhin, daß sowohl aktivierende, als auch reprimierende Elemente in der *DDX3X* Promoterregion enthalten sind (Abb. 4-59). Der Aktivitätsanstieg, von *DDX3X* Exon1 ausgehend, der bis zum Konstrukt VI zu beobachten ist, zeigt deutlich, daß in dieser Region mehrere Enhancer-Elemente lokalisiert sind, die in Konstrukt Nr. VI gemeinsam zu der starken Luziferaseaktivität führen. Der Abfall der *Luc*-Aktivität in Konstrukt VII könnte entweder auf ein Repressor-Element hinweisen, oder auf eine fehlerhafte räumliche Struktur der Promoterregion. Distal von Konstrukt VI scheinen Repressor-Elemente lokalisiert zu sein, die zu einer niedrigeren *Luc*-Aktivität der Konstrukte III - V führen. Zwischen Konstrukt II und III ist wiederum ein Enhancerelement lokalisiert, welches zu einem erneuten Aktivitätsanstieg von Konstrukt II im Reportergen-Assay führt.

Die isolierte Betrachtung ausgewählter *DDX3X* Promoterabschnitte ergab (Abb. 4-60), daß es sich praktisch bei allen getesteten Inserts um Enhancer/Repressor-Elemente handelt, da keine der Sequenzen eine nennenswerte Eigenaktivität im promoterlosen pGl4.10 Vektor zeigte. Einzig Insert h (CpG2) wies eine geringe *Luc*-Eigenaktivität von ~ 17% auf. Die Inserts a, b, d, e und f bestätigen die Ergebnisse der ersten Reportergenserie und zeigen deutliche Enhancer-qualitäten (Abb. 4-61). Insert a (AG-Box) fällt dabei mit einer über 5fachen Aktivitätssteigerung im pGL4.13 Vektor auf. Im Bereich von Insert a ist die X-spezifische AG-reiche Sequenz stromaufwärts zur *DDX3X* T/B-TSS-II Startstelle, lokalisiert. Diese Promoterregion stellt somit ein starkes Enhancer-Element dar. Die Deletion von *DDX3X* Promoterabschnitten im Konstrukt IIΔ zeigt zudem deutlich, daß das Fehlen wichtiger Abschnitte zu einer starken Aktivitätsreduktion führt (Abb. 4-63). Nur zwei der getesteten zwölf Inserts (Insert g und k) zeigten in nativer Orientierung einen Repressoreffekt im Reportergenkonstrukt (Abb. 4-62). Damit wird ein deutlicher Repressoreffekt in die distale Promoterregion k karitert. Die distalen Repressor-Elemente könnten somit zu einer Repression der distalen hodenspezifischen Transkriptions-startstelle T-TSS-III in somatischen Geweben (Soma, sowie Sertoli- und Leydig-Zellen) führen. Es soll aber auch erwähnt werden, daß die zwei Subklone von Insert k, Insert j und l, keinen Repressoreffekt gezeigt haben. Zusammen mit der Feststellung, daß Insert I (Aktivator/Repressor) und c (POZ1) deutliche Abhängigkeiten von der klonierten Polarität zeigten, weisen die Ergebnisse darauf hin, daß die *DDX3X* Promoterabschnitte eine bestimmte Position und einen Sequenzkontext für ihre korrekte Funktion benötigen.
Die Ergebnisse in Abb. 4-63 zeigen also, daß einige 5´ gelegene *DDX3X* Promoterabschnitte direktional unterschiedlich wirken. Wei & Brennan (2000) beschreiben solch eine Subkategorie orientierungs-abhängiger und *cis*-regulativer Elemente, am Beispiel des Drosophila „*adh* larval

enhancer“ (ALE), als „enhancing region“. Andere „Enhancer-Elemente“, die nur bei einer bestimmten Polarität für einen Promoterstimulus sorgen, wurden von Whyte et al. (1999), Swamynathan & Piatigorsky (2002) und Gao et al. (2011) beschrieben. Es besteht somit auch die Möglichkeit, daß die inhärente Funktionalität dieser *DDX3X* Promoterabschnitte, in einem Versuchsaufbau mit pGL4.13 „Fremdpromoter“, nicht korrekt wiedergegeben werden kann. Da die getesteten *DDX3X* Sequenzen innerhalb der funktionellen *DDX3X* Promoterdomäne liegen (Abb. 4-58, 4-60), ist es somit möglich, daß diese *cis*-regulativen Elemente zu der orientierungsabhängigen Subgruppe der „Enhancer-Elemente“ gehören.
Dies könnte auch erklären, warum mit den partiellen *DDX3X* Promotersequenzen (Abb. 4-60) nicht jeder Effekt der durchgehenden Promoterkonstrukte (Abb. 4-58) der ersten „Dual Luciferase Reporter“-Serie reproduziert werden konnte. Generell zeigt sich aber eine gute Übereinstimmung der zwei etablierten *DDX3X* Reporterkonstrukt-Serien. Im proximalen Promoterbereich sind starke Enhancer-Elemente vorhanden und im distalen Bereich sind, neben zusätzlichen Enhancer-Elementen, auch Repressor-Elemente lokalisiert (Abb. 4-59, 4-61 bis 4-63). Das in dieser Arbeit detailliert etablierte *DDX3X* Promoteraktivitätsprofil wurde kürzlich von Botlagunta et al. (2011) durch ein *DDX3X* Aktivitätsprofil, welches mit drei pGL3 Promoter-Reportergenkonstrukten in der weiblichen Brustkrebs-Zellinie MCF 7 ermittelt wurde, bestätigt.

5.5.2. Potentieller Einfluß von G-Quadruplex-Sequenzen auf die *DDX3X* Promoteraktivität

In Kapitel 4.11.2. wurden potentielle G-Quadruplexstrukturen in den *DDX3X* und *DDX3Y* 5´UTR-Sequenzen als mögliche negative Regulatoren der Translationseffizienz identifiziert. Regulative G-Quadruplexe sind nicht nur in 5´- und 3´UTRs von mRNA-Molekülen, sondern auch in DNA-Molekülen, v.a. in Telomer- und Promoterregionen bekannt (Parkinson, 2006; Huppert et al., 2008; Wong & Huppert, 2009; Basundra et al, 2010; Munk, 2010; Wu & Brosh, 2010; Balasubramanian et al., 2011). In den vergangenen zehn Jahren sind regulative G-Quadruplexe in den Fokus der Krebstherapie gerückt, um spezifische G-Quadruplex-bindende Liganden zu finden und zu entwickeln (Amato et al., 2011; Balasubramanian et al., 2011; McLuckie et al., 2011). Die Ionenkonzentration und spezifische Liganden können Bildung und Stabilität intramolekularer G-Quadruplexe beeinflussen, was von großem Interesse für die extrinsische Transkriptionskontrolle bestimmter Proto-Onkogene ist (Balasubramanian et al., 2011).

Im *DDX3X* Promoter sind zusätzlich zu den bereits in Tab. 4-43 beschriebenen RNA G-PQSs (putative quadruplex-forming sequences) auch intramolekulare DNA G-PQSs im Template-Strang möglich, wenn im kodierenden DNA-Strang C-PQSs vorhanden sind (Abb. 5-3). Die Bildung intramolekularer Sekundärstrukturen in doppelsträngigen (C:G)-reichen DNA-Regionen,

ist trotz der höheren Schmelztemperatur, möglich (Balasubramanian et al., 2011). Es ist also denkbar, daß die G-Quadruplexe erst nach Öffnung der DNA-Doppelhelix, z.B. während der Transkription, gebildet und anschließend durch Liganden stabilisiert werden.

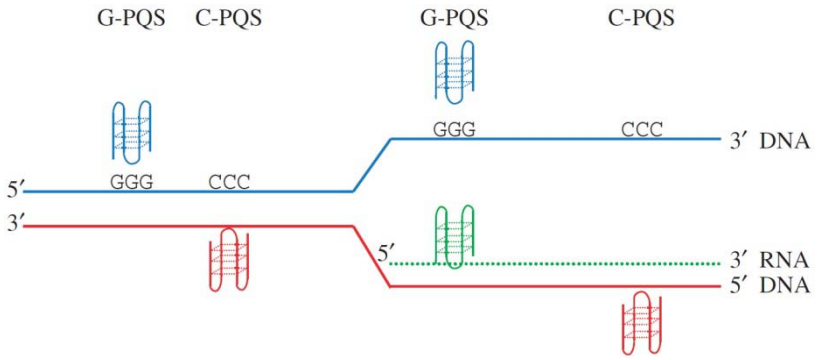


Abb. 5-3: Schematische Darstellung potentieller G-Quadruplexsequenzen in DNA- und RNA-Molekülen.
Kodierender DNA-Strang (codogener Strang) = blau, Templat-Strang (nicht-codogener Strang) = rot, RNA-Strang = grün. PQS = „putative quadruplex sequences“ (nach Huppert et al., 2008).

Innerhalb der Pyrimidin-reichen *DDX3X* Region (BAC-Klon RP13-13A3, Pos. 29.816 - 29.978), im Bereich der T-TSS-III Position (BAC-Pos. 29.825), ist dies möglich. Von BAC-Pos. 29.818 - 29.846 (29 bp = C-PQS1) sind auf dem Templat-Strang mehrere überlappende G-Quadruplexe mit einem maximalen G-Score von 52 möglich, die die *DDX3X* Promoteraktivität beeinflussen könnten, da die hodenspezifische *DDX3X* T-TSS-III Position genau in dieser Sequenz liegt. Anhand der genomischen Sequenz sind mehrere G-Quadruplexkonfigurationen möglich und ist es wahrscheinlich, daß mehrere Strukturen in einem dynamischen Gleichgewicht zueinander stehen (Balasubramanian et al., 2011).
Eine zweite C-PQS Region ist von BAC-Pos. 29.893 - 29.924 (32 bp = C-PQS2), mit einem maximalen G-Score von 47, möglich. Beide Regionen C-PQS1 und C-PQS2 kommen im distalen 5´ Bereich des *DDX3X* Promoterkonstruktes VII vor (Tab. 4-27; Abb. 4-58). In Konstrukt Nr. VIII ist nur C-PQS2 vorhanden. Die geringere Promoteraktivität des Konstruktes VII (67,6%) im Dual-Luziferase Reportergenassay, im Vergleich zu Konstrukt VIII (77,3%), könnte also auf die Bildung dieser G-Quadruplexstrukturen hindeuten, die in Position und Anzahl unterschiedliche Auswirkungen auf die *DDX3X* Promoteraktivität nehmen (Abb. 4-59). Erwähnt werden soll, daß mit dem Insert f (TC-Box, BAC-Pos. 29.785 - 29.981) ein möglicher Repressor-Effekt durch G-Quadruplexe nicht festgestellt werden konnte (Abb. 4-61). Ein Unterschied der Reportergenaktivierung zwischen den Subklonen und den längeren Promotersequenzen ist, wie oben in Kapitel 5.5.1. beschrieben, zu diskutieren.

Im *DDX3Y* Promoter sind am Übergang von Exon-Tb zu Intron-T (BAC-Klon 475I1, Pos. 54.711 - 54.742 = 32 bp) ebenfalls mehrere G-Quadruplexkonfigurationen mit einem maximalen G-Score von 47 möglich. Diese könnten funktionell für die gewebespezifische Zugänglichkeit der

hodenspezifischen *DDX3Y* Core-Promotermotive durch den PIC sein. Als Ursache einiger gewebespezifischer Spliceereignisse, wird die Bildung von G-Quadruplexen vermutet (Kostadinov et al., 2006; Scaria et al., 2006). Dies könnte somit eine Erklärung für die G-PQSs am Übergang von Exon-Tb zu Intron-T sein. Zudem konnte auch gezeigt werden, daß G-PQS (Q1, Q2) in den *MSY2* Repeatsequenzen möglich sind (Abb. 4-72). Minisatelliten tendieren generell dazu alternative DNA-Konformationen, wie z.B. G-Quadruplexe, zu etablieren (Mirkin, 2007; Schonhoft et al., 2009, 2010).

5.5.3. Weiterführende Projekte zur Bestimmung der *DDX3X* Promoterfunktionen

Auf Grund der hier ermittelten Daten zur *DDX3X* Promoterregion können gezielte Folgeexperimente vorgeschlagen werden. Da bekannt ist, daß in unterschiedlichen Zellinien, auf Grund deren Herkunft wie Gewebe und Geschlecht, unterschiedliche TFs und Core-Promoter-Erkennungsfaktoren exprimiert sein können, wird vorgeschlagen die Daten aus der EP2102 Zellinie in weiteren Zellinien wie DU145 und LNCaP (Prostatakarzinom, ♂), HeLa (♀) und HEK 293T ⚥) zu vergleichen. Da bei den hier durchgeführten Luziferaseexperimenten keine parallelen Trankriptanalysen durchgeführt wurden, ist eine eindeutige Beurteilung von Transkriptions- und Translationseffekten nicht möglich. In weiteren Versuchsreihen sollte daher eine zusätzliche Kontrolle der experimentellen mRNAs erfolgen. Hierdurch kann dann z.B. analysiert werden, an welcher TSS die Transkription in den Reportergenkonstrukten beginnt und wie viele der kartierten uATGs (Abb. A-5) in den Reporter-Transkripten tatsächlich vorkommen. Eine Möglichkeit zur Vermeidung artifizieller Effekte der uATGs und uORFs wäre auch eine gezielte Mutation dieser Elemente in den Reportergenkonstrukten.

Mit Konstrukt IIΔ konnte eine funktionelle Aktivatorqualität der deletierten 929 bp langen Sequenz gezeigt werden (Abb. 4-60, 4-63). Die in dieser Region geclustert vorkommenden Faktoren ZF5 und SP1, sowie die konservierte SOX5/30 Bindestelle, bieten gute Ziele für gezielte Mutationen der TFBSs und Knock-outs der TFs, durch z.B. Antisense-Transkripte. Ebenso sollten die sechs potentiellen TF-Module (Abb. 4-49), die sowohl in der *DDX3X*, als auch in der *DDX3Y* Promoterregion identifiziert wurden, auf ihre Funktionalität überprüft werden. Die molekulare Basis dieser Aktivatorfunktion für die *DDX3X* Genexpression ist zudem durch gezielte Mutationen der drei *DDX3X* Core-Promotermodule in den Reportergen-konstrukten analysierbar.

Für die funktionelle Analyse von CpG-Islands und SP1-Bindungsstellen werden häufig *Drosophila* SL2 Schneider-Zellen verwendet, die defizient für den TF SP1 sind (Nielsen et al., 1998). Die potentielle Funktion des *DDX3X* CpG-Islands und der darin kartierten repetitiven SP1-Bindestellen sollte daher in einem solchen *in-vivo* Zellsystem analysiert werden. Durch Co-Transfektion von SP1-Expressionsvektoren und durch *in-vitro* Methylierung der *DDX3X*

Promoterkonstrukte, kann der Einfluß von SP1 und des CpG-Islands zur *DDX3X* Genregulation untersucht werden. SP1 bindet nur an hypomethylierte Promotersequenzen (Clark et al., 1997; Suske, 1999), daher können sowohl die Funktion von SP1, als auch des CpG-Islands in einem Versuchsaufbau analysiert werden. Als alternatives Experiment, zur Analyse des Methylierungsstatus der *DDX3X* Promoterregion in *Drosophila* SL2 Zellen und den aufwendigen Bisulfit-Sequenzierungen, existiert die Möglichkeit einer PCR-basierten DNA-Methylierungsanalyse („Methylation Promoter PCR Kit“ #MP1100, Panomics; siehe auch Douet et al., 2007).

Zusätzlich zu den oben genannten Analysen können EMSA (Electro-mobility shift assay) und ChiP-Experimente (Chromatin Immunopräzipitation; z.B. „ChiP-IT Kits“ von „Active Motif“) durchgeführt werden, um bestimmte TF/DNA-Bindungen zu bestätigen. Um das TF-Bindungsprofil von Keimzellen und z.B. Leukozyten zu vergleichen, besteht die Möglichkeit „Protein/DNA Array Kits“ mit unterschiedlichen TF-Profilen einzusetzen (#MA1210 - #MA1214, Panomics). Hierbei werden definierte Biotin-markierte DNA-Proben mit Zellextrakten inkubiert und nach Lösung der Protein/DNA-Komplexe werden die Proteine auf eine Membran mit definierten TF-Konsensussequenzen aufgetragen. Anhand der Signalpositionen auf dem Array können die vorhandenen TFs identifiziert werden.

5.6. Funktionelle *MSY2* und *MSY2-X* Analysen

Insgesamt hat es den Anschein, daß in der *DDX3Y MSY2* DLR-Assay Serie nur die zwei Inserts m (*MSY2* Repeat 1) und u (*MSY2* Repeat 4; Abb. 4-65), mit je einer *MSY2*-Repeateinheit, einen positiv regulativen Effekt auf die SV40 Promoteraktivität in der EP2102 Zellinie ausüben (Abb. 4-66). Sobald zwei oder mehr *MSY2* Repeateinheit in den pGL4.13 Vektor inseriert werden, ist ein Repressoreffekt festzustellen (Abb. 4-66). Eine Abhängigkeit der *MSY2* Aktivität vom *DDX3Y* Promoterkontextes ist auch in dieser DLR-Serie möglich. Ein Einfluß der *MSY2* Polarität auf die *DDX3Y* Promoterfunktionalität scheint ebenfalls möglich, da Insert p (*MSY2* Repeat 1-4) in negativer Orientierung zu einer vollständigen Inaktivierung des endogenen SV40-Promoters, im pGL4.13 Vektor, führte (Kapitel 4.10.1). Ob die Anwesenheit der *DDX3Y* T-TSS-II Position in den Inserts n bis s einen Effekt auf die SV40 Promoteraktivität ausübt ließ sich in diesem Versuchsaufbau nicht klären. Hierzu müssten in weiteren Experimenten die in Zellkultur synthetisierten Reportergentranskripte auf ihre Transkriptions-initiation analysiert werden, da eine Transkriptinitiation sowohl im SV40 Promoter/Enhancer, als auch in der *DDX3Y MSY2* Region erfolgen könnte.

Es ist nicht bekannt ist, ob die in EP2102 Zellen beobachteten Effekte der *MSY2* Mutationskonstrukte (Abb. 4-68) spezifisch für EP2102 sind, oder ob diese auf die Keimbahn

übertragbar sind. Für VDR sind bereits reprimierende Eigenschaften in der Literatur beschrieben worden (Li & Gardner, 1994; Murayama et al, 1999; Dong et al., 2005). Die Erhöhung der Luziferaseaktivität nach Mutation der VDR-TFBS in Insert m und n (Abb. 4-68), könnte somit ebenfalls auf einen Repressoreffekt von VDR auf die *DDX3Y* Expression hinweisen.
Die CREB1-TFBS kommt durch die *MSY2* Vervierfachung drei Mal in *MSY2-2*, *-3* und *-4* vor und könnte dadurch eine wichtige Funktion in der *MSY2* Enhancer/Repressor-Funktikon einnehmen. Die Mutation der CREB1-TFBS in *MSY2-2* führte zu einer moderaten Erhöhung der Luziferaseaktivität (Abb. 4-68). Dies ist somit ein Hinweis auf eine CREB1 Repressorfunktion in *MSY2*. Von CREB1 und anderen CREB-TFs der bZIP TF-Familie sind positiv wie negativ regulative Effekte in der Literatur beschrieben (Lamph et al., 1990; Song et al., 2007; Nonaka, 2009). In Kapitel 5.3.2. wurde bereits angeführt, daß die CREB-TFBS zu den Methylierungs-sensitiven TFBS zählt und dadurch eine zusätzliche Spezifität an diesen Bindestellen erzielt werden kann.

Die *MSY2-X* Sequenz, stromaufwärts der *DDX3X* T-TSS-III Startstelle (Abb. 4-69), zeigt in Insert v eine geringe Steigerung der Luziferaseaktivität im pGL4.13 Vektor (Abb 4-70). Das stromabwärts von *MSY2-X* um 78 bp verlängerte Insert w zeigt im pGL4.13 Vektor eine 2,4fache Aktivitätssteigerung im Vergleich zur Positvkontrolle (Abb 4-70). In diesem Bereich des *DDX3X* Promoter ist somit ein starkes Aktivatorelement kartiert, welches auch durch Insert e bestätigt wird (Abb. 4-61). Die kartierten TF-Module SOX-PAX und SP1-SPI1 (Abb. 4-49, siehe Kapitel 5.3.1.), sowie ein Teil der repetitiven ZF5 TFBSs, sind in dieser Schnittmenge der Inserts e und w lokalisiert. Somit konnte zwischen der *MSY2-X* Sequenz und der hodenspezifischen *DDX3X* T-TSS-III Transkriptionstartstelle ein starkes Aktivatorelement für die *DDX3X* Genexpression kartiert werden. Die Reduktion der Luziferaseaktivität nach Mutation der SOX5-TFBS in Insert w zeigt (Abb. 4-71), daß die SOX5 TFBS sehr wahrscheinlich ein funktionelles Element der hodenspezifischen *DDX3X* Genkontrolle ist.

Diese *MSY2* und *MSY2-X* Experimente demonstrieren somit, daß die Kombination aus vergleichender Genomik und gezielten Funktionsanalysen, eine erfolgreiche Strategie zur Identifizierung *cis*-regulativer Elemente in den *DDX3Y* und *DDX3X* Promoterregionen darstellt.

5.7. Funktionelle *DDX3X* 5´UTR Analyse

Die experimentellen Human *DDX3X* 5´UTR *in-vitro* Translationsassays bestätigen im wesentlichen das in Kapitel 4.11.3. (Abb. 4-77) vorgestellte CART-Modell auf seine Vorhersagequalität für die Translationseffizienz der *DDX3X* 5´UTR-Varianten. Die „Klasse I“-

Transkripte mit Start in der *DDX3X* T-TSS-III Position zeigten *in-vitro*, wie vorhergesagt, kaum Translationspotential. Das „Klasse III"-Transkript *DDX3X* TSS-I wies, in Übereinstimmung mit dem CART-Modell, keine Translationsrepression auf.

Einzig für die generierten *DDX3X* TSS-II Transkripte mit Start in BAC-Position 30.280 (BAC-Klon RP13-13A3) scheint die CART-Vorhersage nicht zu stimmen. Obwohl mit einer 5´UTR-Länge von 450 Nukleotiden und einer freien Energie ΔG von -172 kcal/mol die Kriterien eines „Klasse I"-Transkriptes vorlagen, konnte experimentell keine Repression für diese *DDX3X* 5´UTR-Variante nachgewiesen werden.

5.7.1. *DDX3X* 5´UTR und 5´UTR-Splicing

DDX3X Transkripte mit Start in T-TSS-III (BAC-Klon RP13-13A3, Pos. 29.825) können in bis zu 50% ein 5´UTR-Splicing aufweisen (Zimmer J., unveröffentlichte Resultate). Dieser Mechanismus könnte für die humanspezifische Kontrolle der DDX3X Proteinmenge, in bestimmten Phasen, oder Zelltypen der Spermatogenese, von Bedeutung sein. Die Reduktion der Komplexität der möglichen Sekundärstrukturen und Vermeidung stabiler Duplex-Strukturen in der gesplicten *DDX3X* 5´UTR-Variante H2 könnten zu der verbesserten Translation beitragen (Abb. 4-79). Die räumliche Entfernung des *DDX3X* uORF1 vom Haupt-ATG scheint im Vergleich beider T-TSS-III 5´UTR-Varianten nicht von entscheidender Bedeutung für die Translationseffizienz zu sein, da für die gesplicte 5´UTR eine höhere Aktivität ermittelt werden konnte. In der ungesplicten *DDX3X* 5´UTR ist der uORF1 648 bp und in der gesplicten Variante 222 bp vom Haupt-ATG entfernt. Die Nutzung von Splice-Varianten zur Kontrolle der Proteinmengen stellen wichtige Mechanismen dar, die nicht von der Translation selbst abhängen (Kozak, 2002).

5.7.2. *DDX3X* 5´UTR und uORFs

Calvo et al. (2009) haben vier Kriterien einer 5´UTR definiert, die mit einer erhöhten Repression des Haupt-ATGs einhergehen. Diese sind: 1) ein starker uATG Kontext, 2) eine erhöhte Entfernung des uORFs von der 5´-Cap-Struktur (> 100 nts) und 3) das Vorhandensein mehrerer uORFs in einer 5´UTR. Das vierte Kriterium der evolutionären Konserviertheit eines uORFs trifft besonders auf *DDX3X* uORF1 zu (Tab. 4-45). Die Länge eines uORFs (Durchschnittswert für Human uORF ist mit 48 Nukleotiden angegeben) soll im Vergleich dazu einen geringeren Einfluß auf die Translationsfähigkeit einer 5´UTR haben (Calvo et al., 2009). Der *DDX3X* uORF1 ist mit 78 Nukleotiden Länge deutlich über dem Durchschnittswert. Das humane *DDX3X* T-TSS-III Transkript mit Start in Position 29.825 erfüllt somit alle fünf von Calvo et al. (2009) genannten Kriterien für ein Transkript mit geringer Translationseffizienz, da auch die Position

des uORF1 mit einer Entfernung von 180 nts von der 5´Cap-Struktur zu einer erhöhten Repression beitragen kann.
Die Translations-Reinitiation des Ribosoms 3´ eines uORFs erfolgt meist weniger effizient, da nach erfolgter Translationsinitiation weniger als 50% der Ribosomen weiterhin entlang der 5´UTR scannen und reinitiieren können (Jackson et al., 2010). Für eine effektive Reinitiation downstream eines uORFs sollte der uORF generell nicht länger als 35 Aminosäuren (108 Nukleotide) sein und es sollte mindestens ein Abstand von 50 Nukleotiden zum Haupt-ORF bestehen, da der Komplex aus Ribosomenuntereinheiten und Initiationsfaktoren nach der Termination zerfallen kann. Der Translationsinitiationskomplex benötigt in der Folge eine gewisse Zeit, um wieder Faktoren wie GTP und Met-RNA_i zu rekrutieren und wieder translationsaktiv zu werden (Kozak, 2002; Rajkowitsch et al., 2003; Cvijovic et al., 2007; Jackson et al., 2010).
Der schwache ATG-Kontext des *DDX3X* uORF2 könnte einen Prozeß namens „Leaky Scanning Mechanism“ ermöglichen, mit dem ein uORF übergangen werden kann. In solch einem Fall wird die Translation nicht initiiert und das Ribosom scannt weiter entlang der 5´UTR bis ein ATG mit starkem oder adäquatem Translationskontext zur Translationsinitiation oder Reinitiation identifiziert wird (Kozak, 2002). Als positiver Nebeneffekt gewährt das Übergehen eines uORFs der 40s Ribosomen-Untereinheit mehr Zeit, um wieder ein Met-RNA_i zu akquirieren (Kozak, 2002; Jackson et al., 2010).

Die Variabilität der *DDX3X* uORF1-Längen, innerhalb der untersuchten Säugerspezies (Abb. A-4), läßt vermuten, daß wahrscheinlich kein konserviertes Peptid im *DDX3X* uORF1 kodiert ist. Anstatt könnte der uORF1 als *cis*-regulatives Translationskontrollelement konserviert sein. Kleine uORFs werden zwar tatsächlich häufig translatiert, allerdings werden diese kleinen Peptide wieder sehr schnell abgebaut, so daß die Peptide häufig nicht funktionell sind (Kozak, 2002; Kochetov et al., 2008). Eine Antikörper-Generierung gegen die potentiellen *DDX3X* uORF1 und uORF2 Peptide, könnte zeigen, ob diese tatsächlich translatiert werden und ob diese einen gewebespezifischen Einfluß auf die *DDX3X* Translation haben, da sie nur in Hodengewebe entstehen. *DDX3X* uORF2 scheint auf Grund der geringeren vertikalen Konserviertheit keine funktionell konservierte Bedeutung für die *DDX3X* Translationskontrolle in Keimzellen zu haben. Der schwache ATG-Kontext von uORF2 spiegelt dies wider.

Es bleibt zu klären, welchen Einfluß der *DDX3X* uORF1 *in-vivo* auf die Translationsfähigkeit der *DDX3X* 5´UTR ausübt, da in dem hier durchgeführten *in-vitro* Assay je nach Konstrukt (H6 zu H14) gegenteilige Effekte registriert wurden und die uORF1 Mutation nicht zu dem erwarteten Anstieg der Translationsfähigkeit führte (Abb. 4-79, 4-80). Gegen einen positiven Einfluß von uORF1 auf die Translationseffizienz sprechen allerdings die Konstrukte H6, H5 und H3, die

sukzessive am 5´Ende verkürzt wurden, und jeweils weniger *cis*-Elemente beherbergen. Zwischen H6 und H5 ist eine 2,8fache Aktivitätssteigerung festzustellen, nachdem durch Verkürzung der 5´UTR der uORF1, die QGRS1 und die Adenin-reiche Region entfallen. Es ist ebenfalls eine deutliche Veränderung in der freien Energie ΔG von -270 kcal/mol für H6 zu -196 kcal/mol für H5 festzustellen (Tab. 4-49).
Wird Konstrukt H3 noch in diese Betrachtung einbezogen, verdoppelt sich die rel. Luziferase-aktivität nochmals von H5 zu H3, wobei in Konstrukt H3 zusätzlich noch uORF2 und QGRS2 entfallen. Die freie EnergieΔG verändert sich nur geringf ügig von -196 zu -172 kcal/mol. Die Auswertung dieser drei Konstrukte unterstützt also den Eindruck, daß uORF1 einen negativen Einfluß auf die Translationseffizienz der *DDX3X* 5´UTR besitzt. Eine mögliche reprimierende Funktion des *DDX3X* uORF1 ist zudem in Übereinstimmung mit Daten des homologen uORF5 in der *DDX3Y* 5´UTR. Nach Mutation des uORF5 ATGs konnte in HEK 293T Zellkultur eine höhere Translationsaktivität festgestellt werden (Jaroszynski et al., 2011).

Erkenntnisse, daß die Eliminierung, oder Zugabe *cis*-regulativer Elemente, wie z.B. von uORFs, die Translationsfähigkeit nicht zwangsläufig verbessern müssen (Kozak, 2002, 2005), liefern eine mögliche Erklärung, für die erhaltenen Ergebnisse der *DDX3X* uORF1 Mutationsanalyse. Der Effekt von 5´ gelegenen stabilen Sekundärstrukturen kann so dominant sein, daß ohnehin keine oder nur kaum Translation der mRNA möglich ist (Kozak, 2002, 2005). Zudem scheinen auch uORFs selbst Einfluß auf die mRNA Stabilität nehmen zu können (Iacono et al., 2005). Eine weitere Beobachtung in diesem Kontext wurde von Kozak (2002) gemacht, daß der Einfluß von Sekundärstrukturen *in-vivo* geringer zu sein scheint als *in-vitro*, was durch fehlende stabilisierende *trans*-Faktoren *in-vitro* erklärt wird.

5.7.3. *DDX3X* 5´UTR und der potentielle Einfluß auffälliger Sequenzmotive

Verschiedene Untersuchungen führten zu der Annahme, dass spezifische in „*trans*“ agierende Faktoren an Pyrimidin-reiche Sequenzen binden und so die Translation dieser mRNAs modulieren können (Sawicka et al., 2008). Als mögliches Bindemotiv wurde das repetitive Trinukleotid $(CCU)_n$ beschrieben, an welches das RNA-bindende Protein PTB (Polypyrimidine Tract Binding Protein) binden kann (Sawicka et al., 2008). In der Pyrimidin-reichen Sequenz stromabwärts der *DDX3X* T-TSS-III Translationsstartstelle sind mehrere $(CCU)_n$ Sequenzen lokalisiert. Zudem gibt es Hinweise, daß Pyrimidin-reiche Regionen als IRES-Elemente („Internal Ribosome Entry Site“) fungieren können und PTB als genereller ITAF (IRES *trans*-acting Factor), mit RNA-Sekundärstruktur Remodelling-Funktion, angesehen werden kann (Meerovitch et al., 1991; Sawicka et al., 2008).

Mit den Konstrukten H9 bis H12, mit Transkriptstart innerhalb der Pyrimidin-reichen Region, sollte die Wirkung einer 5´TOP-Sequenz auf die *DDX3X* 5´UTR Translationsfähigkeit untersucht werden. Per se sind Pyrimidin-reiche Sequenzen in 5´UTRs nicht als inhibitorische Elemente bekannt. Es besteht aber die Möglichkeit, daß diese zu besonderen Sekundärstrukturen und durch Duplex-Bildung mit Purin-reichen Regionen der 5´UTR, zu stabilen Haarnadelstrukturen („Stem-loops“) in der mRNA führen. Dies wiederum kann eine Translationsrepression zur Folge haben. Teile der distalen Pyrimidin-reichen *DDX3X* 5´UTR-Sequenz (BAC RP13-13A3, Pos. 28.857 - 29.905) interagieren mit (A,G)-reichen 5´UTR-Sequenzen (BAC-Pos. 30.244 - 30.294) stromaufwärts der T/B-TSS-II Position (BAC-Pos. 30.314). Diese bilden eine vermutlich inhibitorische RNA-Duplexstruktur (Abb. 4-76, siehe Kapitel 4.11.3.).
Die Konstrukte H9 und K10, mit 5´TOP-Sequenz, fallen in ihrem Translationspotential noch schwächer aus, als die nativen Konstrukte H1 und H2. Dies war zu erwarten, da im Retikulozyten-Lysat („TnT-Quick Coupled Transcription/Translation Kit“ basiert darauf) wahrscheinlich ein titrierbarer Repressor, als Bestandteil der allgemeinen Translations-maschinerie vorhanden ist, der selektiv 5´TOP mRNAs inhibiert (Biberman & Meyuhas, 1999).

Bergman et al. (2007) zeigten, daß der Lsm-Proteinkomplex, ein Inhibitor der 3´-5´-Exonukleasen, nicht nur in der 3´UTR, sondern auch an 5´-Poly(A)-Sequenzen binden kann. Die Minimalanforderung hierfür ist eine ununterbrochene Adeninsequenz von zehn Nukleotiden. Dies trifft für die Adenin-reiche Region der *DDX3X* 5´UTR zu. Von PABP (Poly(A)-Binding Protein) ist ebenfalls bekannt, daß es in der eigenen und in der CAT (Chloramphenicol Acetyl-Transferase) 5´UTR binden kann und Translations-inhibitorisch, ohne Reduktion der mRNA Level, wirken kann (Melo et al., 2003). D.h. es besteht auch für Poly(A)-Sequenzen in der 5´UTR ein Zusammenhang zwischen mRNA-Stabilität und Translationseffizienz. Eine weitere Funktion solcher „Polypurin (A)-reicher Sequenzen“ (PARS) kann ebenfalls die Ausbildung von IRES-Strukturen sein (Dorokhov et al., 2002). Für zwei einzelne Cytosin-Nukleotide, die innerhalb der PARS liegen, sind SNPs, mit den Allelen C und A, unter den Referenz-SNP-Nummern „rs12008629“ und „rs12008630“ bekannt (Ning et al., 2001). Dadurch würde jeweils eine längere ununterbrochene Adenin-Sequenz entstehen. Die klinische Relevanz dieser SNPs ist bisher nicht bekannt.

5.7.4. Weiterführende Projekte zur Bestimmung der *DDX3X* 5´UTR Funktionen

Für die Fortführung der *DDX3X* 5´UTR-Funktionsanalyse wird vorgeschlagen ein Zellkultur-modell zu etablieren (siehe Jaroszynski et al., 2011 für *DDX3Y* 5´UTR). Ein experimenteller Ansatz zur funktionellen Überprüfung der Translationsfähigkeit der *DDX3X* Transkripte sollte mit Blick auf mögliche *trans*-Faktoren und die kartierten G-Quadruplexe etabliert werden (Abb. 5-4).

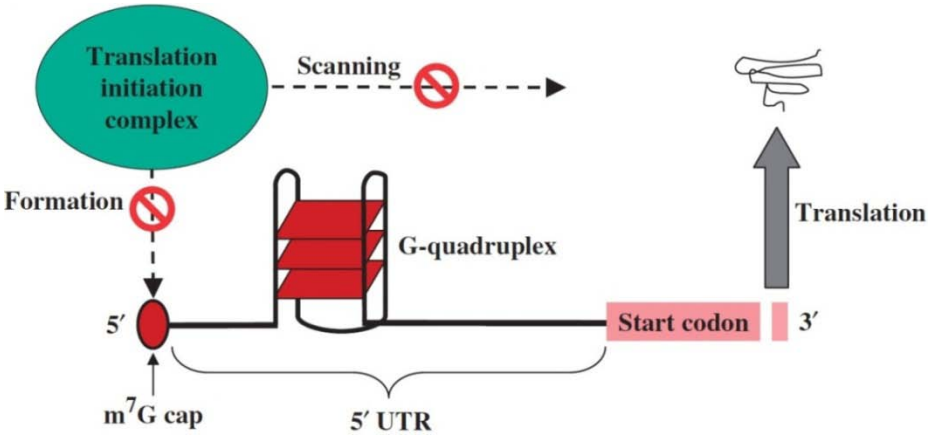


Abb. 5-4: Potentieller Einfluß von G-Quadruplexsequenzen auf die Translationsfähigkeit einer mRNA (nach Huppert et al., 2008).

Kumari et al. (2007, 2008) konnten in Luziferaseassays zeigen, daß die Mutation einer G-Quadruplexsequenz in der 5´UTR des *NRAS* Gens denselben Effekt hatte, wie eine experimentelle Verkürzung der 5´UTR, was ebenfalls zur Entfernung der G-Quadruplex-strukturen führte. Ein solcher Effekt könnte auch für die schrittweise Zunahme der Translationsaktivität der *DDX3X* 5´UTR Konstrukte H1→H5→H3→H4 verantwortlich sein, in denen jeweils weniger G-Quadruplexe vorhanden sind (Abb. 4-79). Gezielte Mutationen der G-Tetraden in den QGRS1 bis -5 sollten zeigen, ob die geringe Translationsfähigkeit von Konstrukt H1 dadurch erhöht werden kann.
Die Beteiligung von G-Quadruplexstrukturen an einer IRES-Struktur (Internal Ribosome Entry Site) konnte ebenfalls gezeigt werden (Morris et al., 2010). Eine experimentelle Überprüfung der Translationsfähigkeit in Zellkultur und *in-vitro* ist also notwendig, um unterschiedliche Eigenschaften der G-Quadruplexe (Aktivierung oder Inhibierung) in beiden Versuchsansätzen vergleichen zu können.

Eine interessante Kontrollmöglichkeit des DDX3 Proteinlevels bietet auch die eigene RNA-Helikaseaktivität der DDX3(X/Y) Proteine (Tarn & Chang, 2009; Lai et al., 2010). So könnten erste DDX3(X/Y) Proteine an die langen strukturierten 5´UTRs der Keimzell-spezifischen *DDX3X* oder *DDX3Y* mRNAs binden, die nur in männlichen Keimzellen vorkommen, und diese effektiver entwinden, als die Helikase eIF4A alleine. Als Endpunkt einer solchen Autoregulation könnte die Inhibierung der Translationsinitiation durch die Komplexbildung von DDX3(X/Y) mit dem 5´Cap-binde Protein eIF4E stehen, welches als kritischer Faktor für die 5´Cap-abhängige Translation gilt (Shih et al., 2008). Möglich ist auch, daß die Translationseffizienz der langen hodenspezifischen Transkripte bewusst gering gehalten werden muß, um schädliche Effekte durch eine zu hohe DDX3(X/Y) Proteinkonzentration in den Keimzellen zu vermeiden.
Es besteht auch die Möglichkeit, daß der Ribosom-Shunt-Mechanismus *in-vivo*, zur Translation der *DDX3X* Transkripte mit langer strukturierter 5´UTR, verwendet wird und *DDX3X in-vivo*

dadurch effizienter als *in-vitro* translatiert wird (Iacono et al., 2005; Morley und Coldwell, 2008). Dabei bindet die kleine Ribosomen-Untereinheit in 5´Cap-abhängiger Weise an die mRNA und überspringt dann weite Bereiche der 5´UTR, um in der Nähe von ATGs den Scanning-mechanismus wieder aufzunehmen (Abb. 1-16). Die *DDX3X* 5´UTR *in-vitro* Translationsassays geben dann nicht mehr die natürliche Translationsfähigkeit dieser langen 5´UTR Transkriptvarianten wieder.

Die hodenspezifische Verlängerung von 5´UTR-Sequenzen, bei gleichzeitiger Verkürzung der 3´UTR, ist ein bekannter Kontrollmechanismus um eine differentielle Genregulation in Keimzellen zu erzeugen (Braun, 1998; Mezquita et al., 1999; Kleene, 2001, 2003; Steger, 2001; Ehrmann & Elliott, 2005; Tian et al., 2005; Zhang et al., 2005; Elliott & Grellscheid, 2006; Davuluri et al., 2008; Lutz, 2008; Sandberg et al., 2008). Mezquita et al. (1999) haben für das „Carbonic Anhydrase II“ Gen (*Car2*) eine somatische und eine hodenspezifische Transkriptvariante in Maus identifiziert. Die hodenspezifische Variante hat vergleichbar zu den hodenspezifischen Human *DDX3X* Transkripten eine verlängerte 5´UTR und eine verkürzte 3´UTR. Entsprechend den hodenspezifischen *DDX3X* Transkripten mit verlängerter 5´UTR, werden *DDX3Y* Transkripte mit Exon-T ebenfalls an einer der proximalen PASs (PAS1, PAS2) in der 3´UTR prozessiert (Rauschendorf et al., 2011).
Entsprechende Transkriptvarianten weisen *in-vitro* und *in-vivo* häufig ein geringes Translationspotential auf (Mezquita et al., 1999). Für das hodenspezifische *Car2* Transkript wurden mehrere Punkte diskutiert, die zu einer verzögerten Translation führen könnten: a) Anwesenheit von Hairpin-Strukturen in der verlängerten 5´UTR, b) potentielle Bindestellen für Translationsregulatoren in der verlängerten 5´UTR und c) Fehlen von destabilisierenden Sequenzen in der verkürzten 3´UTR (Mezquita et al., 1999). Es ist bekannt, daß miRNA-Bindestellen häufig am distalen 3´Ende von 3´UTRs lokalisiert sind, so daß eine gewebeselektive Verkürzung der 3´UTR durch eine interne PAS, tatsächlich zu einer Vermeidung von miRNA-bedingten Abbauprozessen beiträgt (Forman & Coller, 2010; Lutz & Moreira, 2011). *DDX3X* Transkripte mit Start in T-TSS-III könnten dementsprechend bereits in den Transkriptions-aktiven Spermatogonien synthetisiert werden und in einem inaktiven Zustand in mRNPs konzentriert und gespeichert werden. Die *DDX3X* Translationsrepression in Spermatogonien und Reaktivierung in Spermatiden stellt somit wahrscheinlich, zusammen mit der *DDX3Y* Translationskontrolle (aktiv in Spermatogonien, inaktiv in Spermatiden), eine effiziente Regulation der DDX3 Proteinlevel dar, wodurch eine durchgehende Verfügbarkeit eines der beiden DDX3 Proteine, während der Spermatogenese und Spermiogenese, gewährleistet ist.

5.8. Ausblick

Auf Grund der hier ermittelten *DDX3X* und *DDX3Y* RNA-Expressionsmuster wird der Weißbüschelaffe *Callithrix jacchus* als geeignetes Modelltier zum weiteren Studium der *DDX3X/DDX3Y* Keimbahnfunktion vorgeschlagen. Untersuchungen zur vergleichenden Histon- und Protaminexpression in Human und *Callithrix* haben bereits gezeigt, daß *Callithrix* ein geeignetes Modelltier zum Studium der Human Spermatogenese und Spermiogenese sein kann (Saunders et al., 1996; Hecht et al., 2009; Nanassy et al., 2010). Die Keimzellorganisation innerhalb der Hodentubuli ist zudem vergleichbar helikal angeordnet (Millar et al., 2000; Luetjens et al., 2005). Zu klären ist allerdings, ob die vergleichbare *DDX3X* und *DDX3Y* Transkriptionskontrolle in Primaten auch zu einer vergleichbaren Translationskontrolle mit gleicher Verteilung der DDX3X und DDX3Y Proteine führt, oder ob der Humanbefund mit differentieller Proteinexpression vor und nach der Meiose in den Keimzellen humanspezifisch ist.
Die Tatsache, daß für die genomische Rind *Ddx3y* Sequenz im Bereich von Exon-T eine deutliche Homologie und eine hodenspezifische Rind *Ddx3y* Exon-T Expression festgestellt werden konnten, macht das Rind, im Vergleich zur Maus, ebenfalls zu einem geeigneteren Modellorganismus zur Untersuchung der *Ddx3y* Exon-T Funktion. Das Rind ist zudem, auf Grund der praktisch kostenfreien und ethisch unbedenklichen Gewebegewinnung im Schlachthof, ein interessantes Modelltier.

Weiterhin ungeklärt ist die Frage nach der exakten Lokalisation der unterschiedlichen *DDX3X* und *DDX3Y* Transkriptvarianten in den verschiedenen Keimzellen des Human Keimepithels (Übersichtsartikel zum Keimepithel siehe Bergmann, 2005 und Brehm & Steger, 2005). Erste Arbeiten die Transkriptvarianten bestimmten Keimzelltypen zuzuordnen wurden von Jakob (2005; RNA *in-situ* Hybridisierung) und Jaroszynski et al. (2011; Laser-Assisted Capture Procedure, LACP) beschrieben. Alternativ zur LACP haben Gassei et al. (2010) eine effektive Keimzellsortierung per „magnetic-activated cell sorting“ (MACS), mit GFRα1 (Glial cell-derived neutrophic factor family receptor alpha-1) Antikörpern, für Human und *Macaca* Hodengewebe demonstriert. Mit Hilfe einer solchen Keimzellsortierung mit dem membranständigen Protein GFRα1, könnten *DDX3Y* und *DDX3X* Transkriptvarianten nachfolgend spezifisch prä- und postmeiotischen Keimzellenfraktionen zugeordnet werden.

Ein Ansatz zur Klärung welche Transkriptvarianten überhaupt effizient translatiert werden können, sollte durch Polysomenanalyse in einem Dichtegradienten, z.B. über einen Sucrosegradienten (siehe hierzu Kleene et al., 1996; Kleene et al., 2010; Mašek et al., 2011), möglich sein. Nur Polysomen-gebundene mRNAs sollten demnach translatiert werden. mRNA-

Spezies, die an mRNPs gebunden sind, sollten translationell inaktiv sein. Einschränkungen findet diese Methode allerdings, wenn für ein Transkript sowohl mRNP, als auch Polysomen-Fraktionen parallel existieren. Zu klären ist auch, ob jeweils nur eine bestimmte *DDX3X* und *DDX3Y* Transkriptvariante über ein Translationspotential verfügt, oder ob mehrere Transkriptvarianten existieren, die in verschiedenen Zelltypen effektiv translatiert werden können.

Da Polysomenanalysen in Human, auf Grund der sehr geringen Gewebeverfügbarkeit und geringen Anzahl an Keimzellen, praktisch nicht durchführbar sind, bieten, die oben vorgeschlagenen Modelltiere *Callithrix* und Rind, Möglichkeiten solch Material-intensive Experimente dennoch durchführen zu können. Interessant zu sehen wäre eine Verschiebung der Polysomen-gebundenen *DDX3X* und *DDX3Y* Transkriptvarianten vor und nach der Meiose.

Auch sollte im Rahmen der weiteren *DDX3*(*X/Y*) Analyse geklärt werden, ob die Human *DDX3* Pseudogene transkribiert werden (Kim et al., 2001). Für das autosomale Rind *PL10*-Pseudogen wird eine regulative Funktion als Antisense-Transkript diskutiert (Chang & Liu, 2010; Pink et al., 2011).

5.8.1. Diagnostische Möglichkeiten der *DDX3Y* Analyse

Die molekulargenetische Diagnose *cis*-regulativer Genregionen gehört nach wie vor nicht zur Routinediagnostik, da sie arbeits- und zeitaufwendig ist (de Vooght et al., 2009). Ein Entscheidungsbaum für solche molekulardiagnostischen Promoteranalysen wurde von de Vooght et al. (2009) vorgestellt. Die einzelnen Punkte des Entscheidungsbaum, wie Mutation bekannt, Mutation funktionell beschrieben, TFBS betroffen, konservierte TFBS, Aktivitätsveränderung im Luziferaseassay etc, sollen jeweils mit ja, oder nein beantwortet werden. Eine Fortführung und weitere Charakterisierung der Sequenzveränderung im Promoter soll jweils nur bei Nein-Antworten erfolgen, um Resourcen zu sparen. Ein solcher standardisierter Entscheidungsbaum für Promoteranalysen könnte dazu beitragen, neben der Analyse kodierender Genbereiche, die Analyse *cis*-regulativer Sequenzen in der Diagnostik zu etablieren.

Zur Verbesserung der AZF-Gen Mikrodeletionsdiagnostik und Beratung von Patienten mit genetisch bedingter Infertilität, wird die Entwicklung eines DNA-Chips, auch DNA-Microarray genannt, als zusätzliches diagnostisches Instrument vorgeschlagen. DNA-Proben (z.B. DNA-Klone oder PCR-Produkte) werden bei der DNA-Chip Methode auf einer gegitterten Glas- oder Kunststoffoberfläche, in einem zweidimensionalen Koordinatennetz, fixiert (Passarge & Kohlhase, 2006). Die Patienten-DNA wird, nach Isolation aus peripherem Blut, Fluoreszenz-markiert und in Lösung (z.B. Formamid) mit der vorbereiteten DNA auf dem Chip hybridisiert. Die Hybridisierung und Stabilität der DNA-Duplexe ist abhängig von der Anzahl der

Fehlpaarungen. Die Auswertung des Hybridisierungsmusters erfolgt schließlich in einem Microarray-Scanner. Eine fehlende Hybridisierung, auf Grund einer Mutation in der Patienten-DNA, kann folglich durch eine zu geringe Fluoreszenzintensität, im Vergleich zu einer Positivkontrolle, erkannt werden (Passarge & Kohlhase, 2006). Im Vergleich zu klassischen Hybridisierungsmethoden wie der FISH (Fluoreszenz *in-situ* Hybridisierung) und dem Southern Blot, die große Deletionen und Duplikationen erkennen, bietet die DNA-Chip Technologie eine schnelle Methode zur Detektion kleiner Veränderungen. Einer der größten Vorteile der DNA-Chip Technologie ist sicherlich der sehr geringe Einsatz von DNA. Vergleichbar zur PCR-basierten Yq-Mikrodeletionsdiagnostik wäre eine Blutprobe für eine Diagnostik ausreichend. Zudem ermöglich ein DNA-Chip die parallele Analyse vieler Proben (Passarge & Kohlhase, 2006).

Mit einem solchen *DDX3Y* DNA-Chip könnte geklärt werden, ob *DDX3Y* Core-Promotermutationen, oder generell *cis*-regulative *DDX3Y* Mutationen bei infertilen Patienten existieren. Die Qualität der Patientenberatung im Rahmen von ART („assisted reproductive technologies“) würde dadurch erhöht, da *DDX3Y* als Hauptgen im AZFa-Lokus gilt (Vogt et al., 2008; Tyler-Smith & Krausz, 2009). Es besteht dabei die Möglichkeit, daß eine Patientengruppe mit SCO-Pathologie, aber ohne AZF-Mikrodeletion, Mutationen in der *DDX3Y* Promoterregion aufweist und bisher diagnostisch nicht erfaßt wird. Die reproduktionsbiologische Relevanz der *MSY2*-Repeat Vervierfachung im Menschen ist bisher nicht geklärt, jedoch weisen die in dieser Arbeit vorgestellten Ergebnisse deutlich darauf hin, daß der *MSY2* Minisatellitenrepeat ein *cis*-regulatives Elemente der *DDX3Y* Expressions-kontrolle darstellt. Im Menschen konnte die Initiation im Bereich von *MSY2* Repeat 2/3 lokalisiert werden (Abb. 4-8). Bao et al. (2000) zeigten, daß bestimmte chinesische Bevölkerungsgruppen nur über drei *MSY2*-Repeateinheiten verfügen. Die Fertilität der Träger scheint nicht reduziert, da sich dieses Variante in der Bevölkerung stabilisieren konnte. Eine weitere Reduzierung der Repeateinheiten könnte aber zum Wegfall oder zur Verschiebung der Keimzell-spezifischen *DDX3Y* (T-TSS-II, T-TSS-I) Transkriptinitiation führen. Das Patientenscreening sollte folglich klären, ob bei SCO-Patienten Polymorphismen mit weniger als drei *MSY2*-Kopien existieren. Der nicht-kodierende $(TGG)_4$ TNR, zwischen *DDX3Y* TSS-I-(T) und TSS-Ia, für den bereits ein SNP (Referenz-SNP-Nummer „rs13304111“) bekannt ist, sollte mit dem *DDX3Y* DNA-Chip ebenfalls auf eine Assoziation zu genetisch bedingter Infertilität untersucht werden. Der Verlust der *DDX3Y* Exon-T Expression scheint mit dem Verlust von Keimzellen und einem SCO-Syndrom in Verbindung zu stehen (Jaroszynski et al., 2011). Ob dies jedoch kausal eine der möglichen Ursachen für ein SCO-Syndrom darstellt, ist bisher weiter unbekannt. Die Analyse *cis*-regulativer Sequenzen im *DDX3Y* Promoter, eines der wichtigsten AZF-Gene mittels DNA-Chip, sollte jedoch dazu beitragen, die Diagnose eines echten SCO-Syndroms qualitativ verbessern.

6. Literaturverzeichnis

(International Human Genome Sequencing Consortium), Abdellah Z, Ahmadi A, Ahmed S, Aimable M, Ainscough R, Almeida J, Almond C, Ambler A, Ambrose K, Ambrose K et al. (2004). Finishing the euchromatic sequence of the human genome. *Nature*. 431(7011):931-45.

Adkins NL, Hagerman TA, Georgel P. (2006). GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol*. 84(4):559-67.

Affara NA. (2001). The role of the Y chromosome in male infertility. *Expert Rev Mol Med*. 2001:1-16.

Agate RJ, Choe M, Arnold AP. (2004). Sex differences in structure and expression of the sex chromosome genes CHD1Z and CHD1W in zebra finches. *Mol Biol Evol*. 21(2):384-96.

Aguilar V, Fajas L. (2010). Cycling through metabolism. *EMBO Mol Med*. 2(9):338-48.

Ahmad KF, Engel CK, Privé GG. (1998). Crystal structure of the BTB domain from PLZF. *Proc Natl Acad Sci USA*. 95(21):12123-8.

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. (1994). *Molecular Biology of the Cell*. Garland Science, New York, USA, 3. Auflage.

Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. (2010). Annotating non-coding regions of the genome. *Nat Rev Genet*. 11(8):559-71.

Ali S, Wallace RB. (1988). Intrinsic polymorphism of variable number tandem repeat loci in the human genome. *Nucleic Acids Res*. 16(17):8487-96.

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev*. 14(5):521-35.

Amati B, Land H. (1994). Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr Opin Genet Dev*. 4(1):102-8.

Amato J, Pagano B, Borbone N, Oliviero G, Gabelica V, Pauw ED, D'Errico S, Piccialli V, Varra M, Giancola C, et al. (2011). Targeting G-Quadruplex Structure in the Human c-Kit Promoter with Short PNA Sequences. *Bioconjug Chem*. 22(4):654-663.

Andrés O, Kellermann T, López-Giráldez F, Rozas J, Domingo-Roura X, Bosch M. (2008). RPS4Y gene family evolution in primates. *BMC Evol Biol*. 8:142.

Andrews PW, Goodfellow PN, Shevinsky LH, Bronson DL, Knowles BB. (1982). Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. *Int J Cancer*. 29(5):523-31.

Anish R, Hossain MB, Jacobson RH, Takada S. (2009). Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. *PLoS One*. 4(4):e5103.

Antequera F. (2003). Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci*. 60(8):1647-58.

Aquila S, Guido C, Perrotta I, Tripepi S, Nastro A, Andò S. (2008). Human sperm anatomy: ultrastructural localization of 1alpha,25-dihydroxyvitamin D receptor and its possible role in the human male gamete. *J Anat*. 213(5):555-64.

Arkov AL, Ramos A. (2010). Building RNA-protein granules: insight from the germline. *Trends Cell Biol*. 20(8):482-90.

Armstrong SJ, Kirkham AJ, Hultén MA. (1994). XY chromosome behaviour in the germ-line of the human male: a FISH analysis of spatial orientation, chromatin condensation and pairing. *Chromosome Res*. 2(6):445-52.

Asahara H, Santoso B, Guzman E, Du K, Cole PA, Davidson I, Montminy M. (2001). Chromatin-dependent cooperativity between constitutive and inducible activation domains in CREB. *Mol Cell Biol*. 21(23):7892-900.

Atchison L, Ghias A, Wilkinson F, Bonini N, Atchison ML. (2003). Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J*. 22(6):1347-58.

Augustin R, Lichtenthaler SF, Greeff M, Hansen J, Wurst W, Trümbach D. (2011). Bioinformatics identification of modules of transcription factor binding sites in Alzheimer's disease-related genes by in silico promoter analysis and microarrays. *Int J Alzheimers Dis*. 2011:154325.

Avni D, Biberman Y, Meyuhas O. (1997). The 5' terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. *Nucleic Acids Res*. 25(5):995-1001.

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science*. 324(5935):1720-3.

Baek D, Davis C, Ewing B, Gordon D, Green P. (2007). Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res*. 17(2):145-55.

Baird SD, Lewis SM, Turcotte M, Holcik M. (2007). A search for structurally similar cellular internal ribosome entry sites. *Nucleic Acids Res*. 35(14):4664-77.

Baird SD, Turcotte M, Korneluk RG, Holcik M. (2006). Searching for IRES. *RNA*. 12(10):1755-85.

Bajic VB, Tan SL, Christoffels A, Schönbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, et al. (2006). Mice and men: their promoter properties. *PLoS Genet*. 2(4):e54.

Balakirev ES, Ayala FJ. (2003). Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet*. 37:123-51.

Balasubramanian S, Hurley LH, Neidle S. (2011). Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat Rev Drug Discov. 10(4):261-75.

Bao W, Zhu S, Pandya A, Zerjal T, Xu J, Shu Q, Du R, Yang H, Tyler-Smith C. (2000). *MSY2*: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. *Gene*. 244(1-2):29-33.

Bardwell VJ, Treisman R. (1994). The POZ domain: a conserved protein-protein interaction motif. *Genes Dev*. 8(14):1664-77.

Barendziak T. (2006). Untersuchungen zum keimbahnspezifischen Lamin B3 Protein in Vertebraten. Dissertation, vorgelegt dem Fachbereich für Biologie und Chemie der Universität Bremen.

Barrionuevo F, Bagheri-Fam S, Klattig J, Kist R, Taketo MM, Englert C, Scherer G. (2006). Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod*. 74(1):195-201.

Barrionuevo F, Scherer G. (2010). SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol*. 42(3):433-6.

Bartolomei MS. (2009). Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev*. 23(18):2124-33.

Basundra R, Kumar A, Amrane S, Verma A, Phan AT, Chowdhury S. (2010). A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1. *FEBS J*. 277(20):4254-64.

Batey RT, Rambo RP, Doudna JA. (1999). Tertiary Motifs in RNA Structure and Folding. *Angew Chem Int Ed Engl*. 38(16):2326-2343.

Batzer MA, Deininger PL. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet*. 3(5):370-9.

Batzoglou S. (2003). LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 13(4):721-31.

Beggs AH, Migeon BR. (1989). Chromatin loop structure of the human X chromosome: relevance to X inactivation and CpG clusters. *Mol Cell Biol*. 9(6):2322-31.

Behre G, Smith LT, Tenen DG. (1999). Use of a promoterless Renilla luciferase vector as an internal control plasmid for transient co-transfection assays of Ras-mediated transcription activation. *Biotechniques*. 26(1):24-6, 28.

Beilina A, Tassone F, Schwartz PH, Sahota P, Hagerman PJ. (2004). Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet*. 13(5):543-9.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. (2004). Ultraconserved elements in the human genome. *Science*. 304(5675):1321-5.

Benda C. (1891). Neue Mitteilungen über die Entwicklung der Genitaldrüsen und die Metamorphose der Samenzellen (Histogenese der Spermatozoen). Verhandlungen der Berliner Physiologischen Gesellschafi. *Arch. Anat. Physiol.* 1891:549-552.

Benham CJ. (1996). Computation of DNA structural variability - a new predictor of DNA regulatory regions. *Comput Appl Biosci*. 12(5):375-81.

Benkoussa M, Brand C, Delmotte MH, Formstecher P, Lefebvre P. (2002). Retinoic acid receptors inhibit AP1 activation by regulating extracellular signal-regulated kinase and CBP recruitment to an AP1-responsive promoter. *Mol Cell Biol*. 22(13):4522-34.

Ben-Tabou de-Leon S, Davidson EH. (2007). Gene regulation: gene control network in development. *Annu Rev Biophys Biomol Struct*. 36:191.

Bérard S, Rivals E. (2003). Comparison of minisatellites. *J Comput Biol*. 10(3-4):357-72.

Berckmans B, De Veylder L. (2009). Transcriptional control of the cell cycle. *Curr Opin Plant Biol*. 12(5):599-605.

Bergen AW, Pratt M, Mehlman PT, Goldman D. (1998). Evolution of RPS4Y. *Mol Biol Evol*. 15(11):1412-9.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*. 133(7):1266-76.

Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA, Persson A, Ottosson J, Wernérus H, Nilsson P, et al. (2008). A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteomics*. 7(10):2019-27.

Bergmann M. (2005). Spermatogenese - Physiologie und Pathophysiologie. *Urologe*. 44(10):1131-8.

Bergman N, Moraes KC, Anderson JR, Zaric B, Kambach C, Schneider RJ, Wilusz CJ, Wilusz J. (2007). Lsm proteins bind and stabilize RNAs containing 5' poly(A) tracts. *Nat Struct Mol Biol*. 14(9):824-31.

Bernard V, Brunaud V, Lecharny A. (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics*. 11:166.

Biberman Y, Meyuhas O. (1999). TOP mRNAs are translationally inhibited by a titratable repressor in both wheat germ extract and reticulocyte lysate. *FEBS Lett*. 456(3):357-60.

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. (2007). The delayed rise of present-day mammals. *Nature*. 446(7135):507-12.

Bird A. (2011). The Dinucleotide CG as a Genomic Signalling Module. *J Mol Biol*. 409(1):47-53.

Blackwood EM, Eisenman RN. (1992). Regulation of Myc: Max complex formation and its potential role in cell proliferation. *Tohoku J Exp Med*. 168(2):195-202.

Blair Hedges S, Kumar S. (2003). Genomic clocks and evolutionary timescales. *Trends Genet*. 19(4):200-6.

Blazek E, Mittler G, Meisterernst M. (2005). The mediator of RNA polymerase II. *Chromosoma*. 113(8):399-408.

Blomberg Jensen M, Nielsen JE, Jørgensen A, Rajpert-De Meyts E, Kristensen DM, Jørgensen N, Skakkebaek NE, Juul A, Leffers H. (2010). Vitamin D receptor and vitamin D metabolizing enzymes are expressed in the human male reproductive tract. *Hum Reprod*. 25(5):1303-11.

Blum MG, Jakobsson M. (2011). Deep divergences of human gene trees and models of human origins. *Mol Biol Evol*. 28(2):889-98.

Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C, Kohwi-Shigematsu T. (1992). Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*. 255(5041):195-7.

Bois P, Jeffreys AJ. (1999). Minisatellite instability and germline mutation. *Cell Mol Life Sci*. 55(12):1636-48.

Bondy CA, Cheng C. (2009). Monosomy for the X chromosome. *Chromosome Res*. 17(5):649-58.

Bonifer C, Yannoutsos N, Krüger G, Grosveld F, Sippel AE. (1994). Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucleic Acids Res*. 22(20):4202-10.

Bonnal S, Schaeffer C, Créancier L, Clamens S, Moine H, Prats AC, Vagner S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J Biol Chem*. 278(41):39330-6.

Botlagunta M, Krishnamachary B, Vesuna F, Winnard PT Jr, Bol GM, Patel AH, Raman V. (2011). Expression of DDX3 is directly modulated by hypoxia inducible factor-1 alpha in breast epithelial cells. *PLoS One*. 6(3):e17563.

Boulikas T. (1995). Chromatin domains and prediction of MAR sequences. *Int Rev Cytol*. 162A:279-388.

Bowers HA, Maroney PA, Fairman ME, Kastner B, Lührmann R, Nilsen TW, Jankowsky E. (2006). Discriminatory RNP remodeling by the DEAD-box protein DED1. *RNA*. 12(5):903-12.

Braun RE. (1998). Post-transcriptional control of gene expression during spermatogenesis. *Semin Cell Dev Biol*. 9(4):483-9.

Brehm R, Steger K. (2005). *Regulation of Sertoli Cell and Germ Cell Differentiation*. Buchreihe: Advances in Anatomy, Embryology and Cell Biology. Springer Verlag, Berlin, 1. Auflage.

Brohede J, Rand KN. (2006). Evolutionary evidence suggests that CpG island-associated Alus are frequently unmethylated in human germline. *Hum Genet*. 119(4):457-8.

Brouwer JR, Willemsen R, Oostra BA. (2009). Microsatellite repeat instability and neurological disease. *Bioessays*. 31(1):71-83.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E; NISC Comparative Sequencing Program, Green ED, Sidow A, Buard J, Jeffreys AJ. (1997). Big, bad minisatellites. *Nat Genet*. 15(4):327-8.

Brudno M, Morgenstern B. (2002). Fast and sensitive alignment of large genomic sequences. *Proc IEEE Comput Soc Bioinform Conf*. 1:138-47.

Buard J, Brenner C, Jeffreys AJ. (2002). Evolutionary fate of an unstable human minisatellite deduced from sperm-mutation spectra of individual alleles. *Am J Hum Genet*. 70(4):1038-43.

Bulger M, Groudine M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev Biol*. 339(2):250-7.

Bull JJ. (1983). *Evolution of sex determining mechanisms*. Benjamin/ Cummings Pub. Co., Advanced Book Program, Menlo Park, USA.

Burke TW, Kadonaga JT. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev*. 11(22):3020-31.

Bussemakers MJ, van de Ven WJ, Debruyne FM, Schalken JA. (1991).Identification of high mobility group protein I(Y) as potential progression marker for prostate cancer by differential hybridization analysis. *Cancer Res*. 51(2):606-11.

Bustin M, Reeves R. (1996). High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog Nucleic Acid Res Mol Biol*. 54:35-100.

Cai R, Xu W, Dai B, Cai X, Xu R, Lu J. (2010). SATB1 binds an intronic MAR sequence in human PI3kgamma in vitro. *Mol Biol Rep*. 37(3):1461-5.

Cai S, Han HJ, Kohwi-Shigematsu T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet*. 34(1):42-51.

Caiafa P, Zampieri M. (2005). DNA methylation and chromatin structure: the puzzling CpG islands. *J Cell Biochem*. 94(2):257-65.

Cain CE, Blekhman R, Marioni JC, Gilad Y. (2011). Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics*. 187(4):1225-34.

Cairns BR. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*. 461(7261):193-8.

Calvo SE, Pagliarini DJ, Mootha VK. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA*. 106(18):7507-12.

Cameron KE, Resnik J, Webster NJ. (1992). Transcriptional regulation of the human insulin receptor promoter. *J Biol Chem*. 267(24):17375-83.

Cao YX, Jean JC, Williams MC. (2000). Cytosine methylation of an Sp1 site contributes to organ-specific and cell-specific regulation of expression of the lung epithelial gene t1alpha. *Biochem J*. 350 Pt 3:883-90.

Capell BC, Collins FS. (2006). Human laminopathies: nuclei gone genetically awry. *Nat Rev Genet*. 7(12):940-52.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*. 21:2933-42.

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. (2002). Selection for short introns in highly expressed genes. *Nat Genet*. 31(4):415-8.

Castrillon DH, Quade BJ, Wang TY, Quigley C, Crum CP. (2000). The human VASA gene is specifically expressed in the germ cell lineage. *Proc Natl Acad Sci USA*. 97(17):9585-90.

Cavener DR. (1987). Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. *Nucleic Acids Res*. 15(4):1353-61.

Champ PC, Maurice S, Vargason JM, Camp T, Ho PS. (2004). Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Res*. 32(22):6501-10.

Chang TC, Liu WS. (2010). The molecular evolution of PL10 homologs. *BMC Evol Biol*. 10:127.

Chao CH, Chen CM, Cheng PL, Shih JW, Tsou AP, Lee YH. (2006). DDX3, a DEAD box RNA helicase with tumor growth-suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor. *Cancer Res*. 66(13):6579-88.

Charlesworth B, Langley CH, Stephan W. (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics*. 112(4):947-62.

Charlesworth B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr Biol*. 6(2):149-62.

Charlesworth D, Charlesworth B, Marais G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*. 95(2):118-28.

Chatterjee HJ, Ho SY, Barnes I, Groves C. (2009). Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol*. 9:259.

Chatterjee S, Pal JK. (2009). Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell*. 101(5):251-62.

Chelly J, Concordet JP, Kaplan JC, Kahn A. (1989). Illegitimate transcription: transcription of any gene in any cell type. *Proc Natl Acad Sci USA*. 86(8):2617-21.

Chen HZ, Tsai SY, Leone G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer*. 9(11):785-97.

Chen SJ. (2008). RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annu Rev Biophys*. 37: 197–214.

Chen SJ. (2008). RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annu Rev Biophys*. 37:197–214.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 31(13):3497-500.

Chernov IP, Akopov SB, Nikolaev LG, (2004). Structure and Functions of Nuclear Matrix Associated Regions (S/MARs). Russian Journal of Bioorganic Chemistry. 30(1):1-11.

Chesnokov IN, Schmid CW. (1995). Specific Alu binding protein from human sperm chromatin prevents DNA methylation. *J Biol Chem*. 270(31):18539-42.

Chew LJ, Gallo V. (2009). The Yin and Yang of Sox proteins: Activation and repression in development and disease. *J Neurosci Res*. 87(15):3277-87.

Chiappetta G, Avantaggiato V, Visconti R, Fedele M, Battista S, Trapasso F, Merciai BM, Fidanza V, Giancotti V, Santoro M, Simeone A, Fusco A. (1996). High level expression of the HMGI (Y) gene during embryonic development. *Oncogene*. 13(11):2439-46.

Chiaroni J, Underhill PA, Cavalli-Sforza LL. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA*. 106(48):20174-9.

Choi JK. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biol*. 11(7):R70.

Chuang RY, Weaver PL, Liu Z, Chang TH. (1997). Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science*. 275(5305):1468-71.

Chung, J., Lee, S.-G., Song, K. Identification of a human homolog of a putative RNA helicase gene (mDEAD3) expressed in mouse erythroid cells. *Korean J. Biochem*. 27: 193-197, 1995.

Churbanov A, Winters-Hilt S, Koonin EV, Rogozin IB. (2008). Accumulation of GC donor splice signals in mammals. *Biol Direct*. 3:30.

Clark SJ, Harrison J, Molloy PL. (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene*. 195(1):67-71.

Cler E, Papai G, Schultz P, Davidson I. (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci*. 66(13):2123-34.

Cockerill PN, Garrard WT. (1986a). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell*. 44(2):273-82.

Cockerill PN, Garrard WT. (1986b). Chromosomal loop anchorage sites appear to be evolutionarily conserved. *FEBS Lett*. 204(1):5-7.

Cockerill PN, Yuen MH, Garrard WT. (1987). The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J Biol Chem*. 262(11):5394-7.

Cohen DR, Vandermark SE, McGovern JD, Bradley MP. (1993). Transcriptional regulation in the testis: a role for transcription factor AP-1 complexes at various stages of spermatogenesis. *Oncogene*. 8(2):443-55.

Collins FS, Morgan M, Patrinos A. (2003). The Human Genome Project: lessons from large-scale biology. *Science*. 300(5617):286-90.

Collins T, Stone JR, Williams AJ. (2001). All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol Cell Biol*. 21(11):3609-15.

Connor F, Cary PD, Read CM, Preston NS, Driscoll PC, Denny P, Crane-Robinson C, Ashworth A. (1994). DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res*. 22(16):3339-46.

Corà D, Di Cunto F, Provero P, Silengo L, Caselle M. (2004). Computational identification of transcription factor binding sites by functional analysis of sets of genes sharing overrepresented upstream motifs. *BMC Bioinformatics*. 5:57.

Corà D, Herrmann C, Dieterich C, Di Cunto F, Provero P, Caselle M. (2005). Ab initio identification of putative human transcription factor binding sites by comparative genomics. *BMC Bioinformatics*. 6:110.

Corbett ST, Hill O, Nangia AK. (2006). Vitamin D receptor found in human sperm. *Urology*. 68(6):1345-9.

Cordaux R, Batzer MA. (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 10(10):691-703.

Cordin O, Banroques J, Tanner NK, Linder P. (2006). The DEAD-box protein family of RNA helicases. *Gene*. 367:17-37.

Costoya JA. (2007). Functional analysis of the role of POK transcriptional repressors. *Brief Funct Genomic Proteomic*. 6(1):8-18.

Coutinho HD, Falcão-Silva VS, Gonçalves GF, da Nóbrega RB. (2009). Molecular ageing in progeroid syndromes: Hutchinson-Gilford progeria syndrome as a model. *Immun Ageing*. 6:4.

Cremer T, Cremer C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*. 2(4):292-301.

Cross SH, Bird AP. (1995). CpG islands and genes. *Curr Opin Genet Dev*. 5(3):309-14.

Cummings CJ, Zoghbi HY. (2000). Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet*. 1:281-328.

Cvijović M, Dalevi D, Bilsland E, Kemp GJ, Sunnerhagen P. (2007). Identification of putative regulatory upstream ORFs in the yeast genome using heuristics and evolutionary conservation. *BMC Bioinformatics*. 8:295.

Dadoune JP, Siffroi JP, Alfonsi MF. (2004). Transcription in haploid male germ cells. *Int Rev Cytol*. 237:1-56.

Dai B, Ying L, Cai R, Li Y, Zhang X, Lu J, Qian G. (2006). Identification of a nuclear matrix attachment region like sequence in the last intron of PI3Kgamma. *Biochem Biophys Res Commun*. 341(2):583-90.

D'Alessio JA, Wright KJ, Tjian R. (2009). Shifting players and paradigms in cell-specific transcription. *Mol Cell*. 36(6):924-31.

Dang Q, Auten J, Plavec I. (2000). Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. *J Virol*. 74(6):2671-8.

Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*. 107(4):489-99.

Darvish H, Nabi MO, Firouzabadi SG, Karimlou M, Heidari A, Najmabadi H, Ohadi M. (2011). Exceptional human core promoter nucleotide compositions. *Gene*. 475(2):79-86.

Davidson EH. (2006). *The Regulatory Genome. Gene Regulatory Networks In Development And Evolution*. Academic Press, Burlington, USA, 2. Auflage.

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. (2008). The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 24(4):167-77.

Davuluri RV, Suzuki Y, Sugano S, Zhang MQ. (2000). CART classification of human 5' UTR sequences. *Genome Res*. 10(11):1807-16.

Day DA, Tuite MF. (1998). Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J. Endocrinol*. 157: 361-371.

De Angioletti M, Lacerra G, Sabato V, Carestia C. (2004). Beta+45 G --> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence. *Br J Haematol*. 124(2):224-31.

de Bonis L. (2001). Evolution: Vom Affen zum Menschen. Teil I: Evolution der Primaten. *COMPACT 1/2001*. Spektrum der Wissenschaft, Heidelberg.

De Cesare D, Fimia GM, Sassone-Corsi P. (1999). Signaling routes to CREM and CREB: plasticity in transcriptional activation. *Trends Biochem Sci*. 24(7):281-5.

de la Luna S, Allen KE, Mason SL, La Thangue NB. (1999). Integration of a growth-suppressing BTB/POZ domain protein with the DP component of the E2F transcription factor. *EMBO J*. 18(1):212-28.

de Moor CH, Meijer H, Lissenden S. (2005). Mechanisms of translational control by the 3' UTR in development and differentiation. *Semin Cell Dev Biol*. 16(1):49-58.

de Vooght KM, van Wijk R, van Solinge WW. (2009). Management of gene promoter mutations in molecular diagnostics. *Clin Chem*. 55(4):698-708.

Deato MD, Marr MT, Sottero T, Inouye C, Hu P, Tjian R. (2008). MyoD targets TAF3/TRF3 to activate myogenin transcription. *Mol Cell*. 32(1):96-105.

Deato MD, Tjian R. (2007). Switching of the core transcription machinery during myogenesis. *Genes Dev*. 21(17):2137-49.

Deato MD, Tjian R. (2008). An unexpected role of TAFs and TRFs in skeletal muscle differentiation: switching core promoter complexes. *Cold Spring Harb Symp Quant Biol*. 73:217-25.

Deaton AM, Bird A. (2011). CpG islands and the regulation of transcription. *Genes Dev*. 25(10):1010-22.

Dehari H, Tchaikovskaya T, Rubashevsky E, Sellers R, Listowsky I. (2009). The proximal promoter governs germ cell-specific expression of the mouse glutathione transferase mGstm5 gene. *Mol Reprod Dev*. 76(4):379-88.

Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodríguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, et al. (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem*. 281(46):35404-12.

Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodríguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, et al. (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem*. 281(46):35404-12.

DeJong J. (2006). Basic mechanisms for the control of germ cell gene expression. *Gene*. 366(1):39-50.

Del Campo M, Mohr S, Jiang Y, Jia H, Jankowsky E, Lambowitz AM. (2009). Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J Mol Biol*. 389(4):674-93.

Delbridge ML, Patel HR, Waters PD, McMillan DA, Marshall Graves JA. (2009).Does the human X contain a third evolutionary block? Origin of genes on human Xp11 and Xq28. *Genome Res*. 19(8):1350-60.

Delgadillo RF, Whittington JE, Parkhurst LK, Parkhurst LJ. (2009). The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. Biochemistry. 48(8):1801-9.

Deng W, Roberts SG. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB.Genes Dev. 19(20):2418-23.

Deng W, Roberts SG. (2006). Core promoter elements recognized by transcription factor IIB. *Biochem Soc Trans*. 34(Pt 6):1051-3.

Deng W, Roberts SG. (2007). TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma*. 116(5):417-29.

Denny P, Swift S, Connor F, Ashworth A. (1992). An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J*. 11(10):3705-12.

Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J, et al. (2007). Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res*. 17(6):746-59.

Deogun JS, Donts R, Komina, O, and Ma F. (2004). RNA Secondary Structure Prediction with Simple Pseudoknots. In YPP Chen (ed.), Second Asia-Pacific Bioinformatics Conference (APBC 2004), January 18-22, Dunedin, New Zealand. Australian Computer Society, Sydney, Australia, pp. 239-246.

DePamphilis ML. (1999). Replication origins in metazoan chromosomes: fact or fiction? *Bioessays*. 21(1):5-16.

Dermitzakis, ET, Reymond A, Antonarakis SE (2005). Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet* 6: 151-157.

Dickemann C. (2004). Molekulare Expressionsanalyse des AZFa-Gens DDX3Y und des X-homologen DDX3X-Gens in der männlichen Keimbahn: Kartierung funktioneller 3´UTR-Elemente. Diplomarbeit, vorgelegt der Fakultät für Chemie und Biowissenschaften der Universität Karlsruhe (TH).

Dickinson LA, Dickinson CD, Kohwi-Shigematsu T. (1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. *J Biol Chem*. 272(17):11463-70.

Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. (1992). A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell*. 70(4):631-45.

Dillon N. (2006). Gene regulation and large-scale chromatin organization in the nucleus. *Chromosome Res*. 14(1):117-26.

DiMartino JF, Hayes P, Saga Y, Lee JS. (1994). A novel initiator/promoter element within the CD45 upstream region. *Int Immunol*. 6(8):1279-83.

DiNapoli L, Capel B. (2008). SRY and the standoff in sex determination. *Mol Endocrinol*. 22(1):1-9.

Ditton HJ, Zimmer J, Kamp C, Rajpert-De Meyts E, Vogt PH. (2004). The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control. *Hum Mol Genet*. 13(19):2333-41.

Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. (2007). Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol Cell Biol*. 27(13):4685-97.

Döhr S, Klingenhoff A, Maier H, Hrabé de Angelis M, Werner T, Schneider R. (2005). Linking disease-associated genes to regulatory networks via promoter organization. *Nucleic Acids Res*. 33(3):864-72.

Dolfini D, Zambelli F, Pavesi G, Mantovani R. (2009). A perspective of promoter architecture from the CCAAT box. *Cell Cycle*. 8(24):4127-37.

Don J, Stelzer G. (2002). The expanding family of CREB/CREM transcription factors that are involved with spermatogenesis. *Mol Cell Endocrinol*. 187(1-2):115-24.

Dong X, Lutz W, Schroeder TM, Bachman LA, Westendorf JJ, Kumar R, Griffin MD. (2005). Regulation of relB in dendritic cells by means of modulated association of vitamin D receptor and histone deacetylase 3 with the promoter. *Proc Natl Acad Sci USA*. 102(44):16007-12.

Dörflinger U, Pscherer A, Moser M, Rümmele P, Schüle R, Buettner R. (1999). Activation of somatostatin receptor II expression by transcription factors MIBP1 and SEF-2 in the murine brain. *Mol Cell Biol*. 19(5):3736-47.

Dorokhov YL, Skulachev MV, Ivanov PA, Zvereva SD, Tjulkina LG, Merits A, Gleba YY, Hohn T, Atabekov JG. (2002). Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry. *Proc Natl Acad Sci USA*. 99(8):5301-6.

Douet V, Heller MB, Le Saux O. (2007). DNA methylation and Sp1 binding determine the tissue-specific transcriptional activity of the mouse Abcc6 promoter. *Biochem Biophys Res Commun*. 354(1):66-71.

Dunn KL, Davie JR. (2003). The many roles of the transcriptional regulator CTCF. *Biochem Cell Biol*. 81(3):161-7.

Dunn KL, Zhao H, Davie JR. (2003). The insulator binding protein CTCF associates with the nuclear matrix. *Exp Cell Res*. 288(1):218-23.

Eddy EM. (1974). Fine structural observations on the form and distribution of nuage in germ cells of the rat. *Anat Rec*. 178(4):731-57.

Edmonds M. (2002). A history of poly A sequences: from formation to factors to function. *Prog Nucleic Acid Res Mol Biol*. 71:285-389.

Ehrmann I, Elliott DJ. (2005). Post-transcriptional control in the male germ line. *Reprod Biomed Online*. 10(1):55-63.

Elfakess R, Dikstein R. (2008). A translation initiation element specific to mRNAs with very short 5'UTR that also regulates transcription. *PLoS One*. 3(8):e3094.

Elgar G, Vavouri T. (2008). Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet*. 24(7):344-52.

Ellegren H, Parsch J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 8(9):689-98.

Ellegren H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet*. 12(3):157-66.

Elliott DJ, Grellscheid SN. (2006). Alternative RNA splicing regulation in the testis. *Reproduction*. 132(6):811-9.

Emerson JJ, Kaessmann H, Betrán E, Long M. (2004). Extensive gene traffic on the mammalian X chromosome. *Science*. 303(5657):537-40.

Emig D, Albrecht M. (2011). Tissue-specific proteins and functional implications. *J Proteome Res*. 10(4):1893-903.

Epstein DJ. (2009). *Cis*-regulatory mutations in human disease. *Brief Funct Genomic Proteomic*. 8(4):310-6.

Eulalio A, Behm-Ansmant I, Izaurralde E. (2007). P bodies: at the crossroads of post-transcriptional pathways. *Nat Rev Mol Cell Biol.* 8(1):9-22.

Evans K, Ott S, Hansen A, Koentges G, Wernisch L. (2007). A comparative study of S/MAR prediction tools. *BMC Bioinformatics*. 8:71.

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH. (2004). A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. *Am J Hum Genet*. 74(1):180-7.

Filipowicz W, Bhattacharyya SN, Sonenberg N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*. 9(2):102-14.

Fimia GM, De Cesare D, Sassone-Corsi P. (2000). A family of LIM-only transcriptional coactivators: tissue-specific expression and selective activation of CREB and CREM. *Mol Cell Biol*. 20(22):8613-22.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. (2011). Ensembl 2011. *Nucleic Acids Res*. 39(Database issue):D800-6.

Foresta C, Moro E, Ferlin A. (2001). Y chromosome microdeletions and alterations of spermatogenesis. *Endocr Rev*. 22(2):226-39.

Forman JJ, Coller HA. (2010). The code within the code: microRNAs target coding regions. *Cell Cycle*. 9(8):1533-41.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 32(Web Server issue):W273-9.

Frazier ME, Johnson GM, Thomassen DG, Oliver CE, Patrinos A. (2003). Realizing the potential of the genome revolution: the genomes to life program. *Science*. 300(5617):290-3.

Frech K, Danescu-Mayer J, Werner T. (1997). A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J Mol Biol*. 270(5):674-87.

Frietze S, Lan X, Jin VX, Farnham PJ. (2010). Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J Biol Chem*. 285(2):1393-403.

Frisch M, Frech K, Klingenhoff A, Cartharius K, Liebich I, Werner T. (2002). In silico prediction of scaffold/matrix attachment regions in large genomic sequences. *Genome Res*. 12(2):349-54.

Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. (2008). A code for transcription initiation in mammalian genomes. *Genome Res*. 18(1):1-12.

Fröhlich N. (2008). Molekulare Funktionsanalyse des *MSY2* Promotormoduls für die Keimbahnexpression von *DDX3X* und *DDX3Y*. Bachelorarbeit, vorgelegt der Fakultät für Biowissenschaften der Ruprecht-Karls-Universität Heidelberg.

Frontini M, Soutoglou E, Argentini M, Bole-Feysot C, Jost B, Scheer E, Tora L. (2005). TAF9b (formerly TAF9L) is a bona fide TAF that has unique and overlapping roles with TAF9. *Mol Cell Biol*. 25(11):4638-49.

Fukumura J, Noguchi E, Sekiguchi T, Nishimoto T. (2003). A temperature-sensitive mutant of the mammalian RNA helicase, DEAD-BOX X isoform, DBX, defective in the transition from G1 to S phase. J Biochem. 134(1):71-82.

Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ. (2008). Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol*. 4(12):728-37.

Gaboli M, Kotsi PA, Gurrieri C, Cattoretti G, Ronchetti S, Cordon-Cardo C, Broxmeyer HE, Hromas R, Pandolfi PP. (2001). Mzf1 controls cell proliferation and tumorigenesis. *Genes Dev*. 15(13):1625-30.

Gal M, Katz T, Ovadia A, Yagil G. (2003). TRACTS: A program to map oligopurine.oligopyrimidine and other binary DNA tracts. *Nucleic Acids Res*. 31(13):3682-5.

Galande S, Purbey PK, Notani D, Kumar PP. (2007). The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev*. 17(5):408-14.

Ganapathi M, Srivastava P, Das Sutar SK, Kumar K, Dasgupta D, Pal Singh G, Brahmachari V, Brahmachari SK. (2005). Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinformatics*. 6:126.

Gao S, Zhang T, Zhou X, Zhao Y, Li Q, Guo Y, Cheng H, Zhou R. (2005).. Molecular cloning, expression of Sox5 and its down-regulation of Dmrt1 transcription in zebrafish. *J Exp Zool B Mol Dev Evol*. 304(5):476-83.

Gao SY, Dai YP, Long X, Zhang FT, Yu J. (2011 May 18. Epub ahead of print). Identification of the VIL2 enhancer in human embryonic kidney cells. *Cell Biol Int*.

Gardiner-Garden M, Frommer M. (1987). CpG islands in vertebrate genomes. *Mol Biol*. 196(2):261-82.

Garee JP, Oesterreich S. (2010). SAFB1's multiple functions in biological control-lots still to be done! *J Cell Biochem*. 109(2):312-9.

Gassei K, Ehmcke J, Dhir R, Schlatt S. (2010). Magnetic activated cell sorting allows isolation of spermatogonia from adult primate testes and reveals distinct GFRa1-positive subpopulations in men. *J Med Primatol*. 39(2):83-91.

Gasser SM, Laemmli UK. (1986a). The organisation of chromatin loops: characterization of a scaffold attachment site. *EMBO J*. 5(3):511-8.

Gasser SM, Laemmli UK. (1986b). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell*. 46(4):521-30.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 31(13):3784-8.

Gat Y, Gornish M, Perlow A, Chakraborty J, Levinger U, Ben-Shlomo I, Pasqualotto F. (2010). Azoospermia and Sertoli-cell-only syndrome: hypoxia in the sperm production site due to impairment in venous drainage of male reproductive system. *Andrologia*. 42(5):314-21.

Gatta R, Mantovani R. (2011). NF-Y affects histone acetylation and H2A.Z deposition in cell cycle promoters. *Epigenetics*. 6(4):526-34.

Gavet O, Pines J. (2010). Activation of cyclin B1-Cdk1 synchronizes events in the nucleus and the cytoplasm at mitosis. *J Cell Biol*. 189(2):247-59.

Gebauer F, Hentze MW. (2004). Molecular mechanisms of translational control. Nat Rev Mol Cell Biol. 5(10):827-35.

Gershenzon NI, Ioshikhes IP. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*. 21(8):1295-300.

Gibbons A. (1997). Y chromosome shows that Adam was African. *Science*. 278(5339):804-5.

Gill DR, Pringle IA, Hyde SC. (2009). Progress and prospects: the design and production of plasmid vectors. *Gene Ther*. 16(2):165-71.

Gill G, Pascal E, Tseng ZH, Tjian R. (1994). A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci USA*. 91(1):192-6.

Gläser B, Grützner F, Willmann U, Stanyon R, Arnold N, Taylor K, Rietschel W, Zeitler S, Toder R, Schempp W. (1998). Simian Y chromosomes: species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm Genome*. 9(3):226-31.

Glover DM, Hames BD. (1995). *DNA cloning 3: A practical approach: Complex Genomes*. Oxford University Press, Oxford, UK, 2nd edition, reprint 2002.

Goetting-Minesky MP, Makova KD. (2006). Mammalian male mutation bias: impacts of generation time and regional variation in substitution rates. *J Mol Evol*. 63(4):537-44.

Gómez-Ferrería MA, Rey-Campos J. (2003). Functional domains of FOXJ2. *J Mol Biol*. 329(4):631-44.

Göndör A, Ohlsson R. (2006). Transcription in the loop. *Nat Genet*. 38(11):1229-30.

Goodrich JA, Tjian R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet*. 11(8):549-58.

Gordon S, Akopyan G, Garban H, Bonavida B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*. 25(8):1125-42.

Goto H, Peng L, Makova KD. (2009). Evolution of X-degenerate Y chromosome genes in greater apes: conservation of gene content in human and gorilla, but not chimpanzee. *J Mol Evol*. 68(2):134-44.

Graham FL, Smiley J, Russell WC, Nairn R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*. 36(1):59-74.

Granadino B, Arias-de-la-Fuente C, Pérez-Sánchez C, Párraga M, López-Fernández LA, del Mazo J, Rey-Campos J. (2000). Fhx (Foxj2) expression is activated during spermatogenesis and very early in embryonic development. *Mech Dev*. 97(1-2):157-60.

Groppo R, Richter JD. (2009). Translational control from head to tail. *Curr Opin Cell Biol*. 21(3):444-51.

Gross P, Oelgeschläger T. (2006). Core promoter-selective RNA polymerase II transcription. *Biochem Soc Symp*. (73):225-36.

Hackenberg M, Previti C, Luque-Escamilla PL, Carpena P, Martínez-Aroza J, Oliver JL. (2006). CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics*. 7:446.

Hackenberg M, Barturen G, Carpena P, Luque-Escamilla PL, Previti C, Oliver JL. (2010). Prediction of CpG-island function: CpG clustering vs. sliding-window methods. *BMC Genomics*. May 26;11:327.

Hamann L, Bayer KU, Jensen K, Harbers K. (1994). Interaction of several related GC-box- and GT-box-binding proteins with the intronic enhancer is required for differential expression of the gb110 gene in embryonal carcinoma cells. *Mol Cell Biol*. 14(9):5786-93.

Hamilton TL, Stoneley M, Spriggs KA, Bushell M. (2006). TOPs and their regulation. *Biochem Soc Trans*. 34(Pt 1):12-6.

Hampf M, Gossen M. (2007). Promoter crosstalk effects on gene expression. *J Mol Biol*. 365(4):911-20.

Han F, Wang Z, Wu F, Liu Z, Huang B, Wang D. (2010). Characterization, phylogeny, alternative splicing and expression of Sox30 gene. *BMC Mol Biol*. 11:98.

Han YJ, de Lanerolle P. (2008). Naturally extended CT . AG repeats increase H-DNA structures and promoter activity in the smooth muscle myosin light chain kinase gene. *Mol Cell Biol*. 28(2):863-72.

Hanstein R. (2004). Molekulare Funktionsanalyse der homologen DEAD-Box RNA-Helikasen, DDX3X und DDX3Y, in Lymphozyten und männlicher Keimbahn. Diplomarbeit, vorgelegt der Fakultät für Biowissenschaften der Ruprecht-Karls-Universität Heidelberg.

Hardison RC. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet*. 16(9):369-72.

Harley VR, Lovell-Badge R, Goodfellow PN. (1994). Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res*. 22(8):1500-1.

Hashimoto S, Suzuki Y, Kasai Y, Morohoshi K, Yamada T, Sese J, Morishita S, Sugano S, Matsushima K. (2004). 5'-end SAGE for the analysis of transcriptional start sites. *Nat Biotechnol*. 22(9):1146-9.

Hawkins SM, Kohwi-Shigematsu T, Skalnik DG. (2001). The matrix attachment region-binding protein SATB1 interacts with multiple elements within the gp91phox promoter and is down-regulated during myeloid differentiation. *J Biol Chem*. 276(48):44472-80.

Hecht N, Behr R, Hild A, Bergmann M, Weidner W, Steger K. (2009). The common marmoset (Callithrix jacchus) as a model for histone and protamine expression during human spermatogenesis. *Hum Reprod*. 24(3):536-45.

Hedrick PW. (2007). Sex: differences in mutation, recombination, selection, gene flow, and genetic drift. *Evolution*. 61(12):2750-71.

Heessen S, Fornerod M. (2007). The inner nuclear envelope as a transcription factor resting place. *EMBO Rep*. 8(10):914-9.

Heintzman ND, Ren B. (2007). The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell Mol Life Sci*. 64(4):386-400.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 39(3):311-8.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 459(7243):108-12.

Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW. (1993). Developmental differences in methylation of human Alu repeats. *Mol Cell Biol*. 13(8):4523-30.

Heng HH, Goetze S, Ye CJ, Liu G, Stevens JB, Bremer SW, Wykes SM, Bode J, Krawetz SA. (2004). Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci*. 117(Pt 7):999-1008.

Hernández G. (2008). Was the initiation of translation in early eukaryotes IRES-driven? *Trends Biochem Sci*. 33(2):58-64.

Hertz GZ, Stormo GD. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*. 15(7-8):563-77.

Hiller MA, Lin TY, Wood C, Fuller MT. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev*. 15(8):1021-30.

Hiller MA, Chen X, Pringle MJ, Suchorolski M, Sancak Y, Viswanathan S, Bolival B, Lin TY, Marino S, Fuller MT. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development*. 131(21):5297-308.

Hjelmeland AB, Schilling SH, Guo X, Quarles D, Wang XF. (2005). Loss of Smad3-mediated negative regulation of Runx2 activity leads to an alteration in cell fate determination. *Mol Cell Biol*. 25(21):9460-8.

Ho Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, Wasserman WW. (2005). oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res*. 33(10):3154-64.

Hoffman EK, Trusko SP, Murphy M, George DL. (1990). An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor. *Proc Natl Acad Sci USA*. 87(7):2705-9.

Holaska JM, Lee KK, Kowalski AK, Wilson KL. (2003). Transcriptional repressor germ cell-less (GCL) and barrier to autointegration factor (BAF) compete for binding to emerin in vitro. *J Biol Chem*. 278(9):6969-75.

Holbrook SR. (2008). Structural principles from large RNAs. *Annu Rev Biophys*. 37:445-64.

Holmes MC, Tjian R. (2000). Promoter-selective properties of the TBP-related factor TRF1. *Science*. 288(5467):867-70.

Hong CS, Park BY, Saint-Jeannet JP. (2007). The function of Dmrt genes in vertebrate development: it is not just about sex. *Dev Biol*. 310(1):1-9.

Hore TA, Deakin JE, Marshall Graves JA. (2008). The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. *PLoS Genet*. 4(8):e1000169.

Hou Y, Zhou X, Liu J, Yuan J, Cheng H, Zhou R. (2010). Nuclear factor-Y (NF-Y) regulates transcription of mouse Dmrt7 gene by binding to tandem CCAAT boxes in its proximal promoter. *Int J Biol Sci*. 6(7):655-64.

Houck CM, Rinehart FP, Schmid CW. (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol*. 132(3):289-306.

Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SK, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C, et al. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*. 463(7280):536-9.

Hughes TA. (2006). Regulation of gene expression by alternative untranslated regions. *Trends Genet*. 22(3):119-22.

Huppert JL, Bugaut A, Kumari S, Balasubramanian S. (2008). G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res*. 36(19):6260-8.

Iacono M, Mignone F, Pesole G. (2005). uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene*. 349:97-105.

Ikeda T, Zhang J, Chano T, Mabuchi A, Fukuda A, Kawaguchi H, Nakamura K, Ikegawa S. (2002). Identification and characterization of the human long form of Sox5 (L-SOX5) gene. *Gene*. 298(1):59-68.

Inoue M, Kamachi Y, Matsunami H, Imada K, Uchikawa M, Kondoh H. (2007). PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development. *Genes Cells*. 12(9):1049-61.

Inoue-Murayama M, Hibino E, Iwatsuki H, Inoue E, Hong KW, Nishida T, Hayasaka I, Ito S, Murayama Y. (2008). Interspecies and intraspecies variations in the serotonin transporter gene intron 3 VNTR in nonhuman primates. *Primates*. 49(2):139-42.

Iseli C, Stevenson BJ, de Souza SJ, Samaia HB, Camargo AA, Buetow KH, Strausberg RL, Simpson AJ, Bucher P, Jongeneel CV. (2002). Long-range heterogeneity at the 3' ends of human mRNAs. *Genome Res*. 12(7):1068-74.

Istrail S, Davidson EH. (2005). Logic functions of the genomic cis-regulatory code. Proc Natl Acad Sci USA. 102(14):4954-9.

Itman C, Loveland KL. (2008). SMAD expression in the testis: an insight into BMP regulation of spermatogenesis. *Dev Dyn*. 237(1):97-111.

Iwashita S, Koyama K, Nakamura Y. (2001). VNTR sequence on human chromosome 11p15 that affects transcriptional activity. *J Hum Genet*. 46(12):717-21.

Jackson JA, Trevino AV, Herzig MC, Herman TS, Woynarowski JM. (2003). Matrix attachment region (MAR) properties and abnormal expansion of AT island minisatellites in FRA16B fragile sites in leukemic CEM cells. *Nucleic Acids Res*. 31(21):6354-64.

Jackson RJ, Hellen CU, Pestova TV. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 11(2):113-27.

Jacox E, Gotea V, Ovcharenko I, Elnitski L. (2010). Tissue-specific and ubiquitous expression patterns from alternative promoters of human genes. *PLoS One*. 5(8):e12274.

Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*. 10:136.

Jakob S. (2005). Untersuchungen zur Lokalisation und Funktion der 5´UTR Transkriptvarianten von DDX3X und DDX3Y in der Human-Spermatogenese. Diplomarbeit, vorgelegt der Fakultät für Biowissenschaften der Ruprecht-Karls-Universität Heidelberg.

Janečka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. (2007). Molecular and genomic data identify the closest living relative of primates. Science. 318(5851):792-4.

Jankowsky E, Bowers H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res*. 34(15):4181-8.

Jankowsky E, Putnam A. (2010). Duplex unwinding with DEAD-box proteins. *Methods Mol Biol*. 587:245-64.

Jarman AP, Higgs DR. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J*. 7(11):3337-44.

Jaroszynski L, Zimmer J, Fietz D, Bergmann M, Kliesch S, Vogt PH. (2011). Translational control of the AZFa gene DDX3Y by 5'UTR exon-T extension. *Int J Androl*. 2010 Jun 16, Epub ahead of print.

Jeffreys AJ, Wilson V, Thein SL. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*. 314(6006):67-73.

Jeffreys AJ, Bois P, Buard J, Collick A, Dubrova Y, Hollies CR, May CA, Murray J, Neil DL, Neumann R, Stead JD, Tamaki K, Yardley J. (1997). Spontaneous and induced minisatellite instability. *Electrophoresis*. 18(9):1501-11.

Jeffreys AJ, Neil DL, Neumann R. (1998). Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J*. 17(14):4147-57.

Jeziorska DM, Jordan KW, Vance KW. (2009). A systems biology approach to understanding cis-regulatory module function. *Semin Cell Dev Biol*. 20(7):856-62.

Ji Q, Luo ZX, Yuan CX, Wible JR, Zhang JP, Georgi JA. (2002).The earliest known eutherian mammal. *Nature*. 416(6883):816-22.

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci USA*. 106(17):7028-33.

Ji Z, Tian B. (2009). Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One*. Dec 23;4(12):e8419.

Jiang C, Pugh BF. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*. 10(3):161-72.

Jiang Q, Matsuzaki Y, Li K, Uitto J. (2006). Transcriptional regulation and characterization of the promoter region of the human ABCC6 gene. *J Invest Dermatol*. 126(2):325-35.

Jiang T, Xing B, Rao J. (2008). Recent developments of biological reporter technology for detecting gene expression. *Biotechnol Genet Eng Rev*. 25:41-75.

Jin VX, Singer GA, Agosto-Pérez FJ, Liyanarachchi S, Davuluri RV. (2006). Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. *BMC Bioinformatics*. 7:114.

Jobling MA, Tyler-Smith C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*. 4(8):598-612.

Johnson JA, Grande JP, Roche PC, Kumar R. (1996). Immunohistochemical detection and distribution of the 1,25-dihydroxyvitamin D3 receptor in rat reproductive tissues. *Histochem Cell Biol*. 105(1):7-15.

Johnstone O, Deuring R, Bock R, Linder P, Fuller MT, Lasko P. (2005). Belle is a Drosophila DEAD-box protein required for viability and in the germ line. *Dev Biol*. 277(1):92-101.

Juven-Gershon T, Hsu JY, Kadonaga JT. (2006a). Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans*. 34(Pt 6):1047-50.

Juven-Gershon T, Cheng S, Kadonaga JT. (2006b). Rational design of a super core promoter that enhances gene expression. *Nat Methods*. 3(11):917-22.

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. (2008a). The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol*. 20(3):253-9.

Juven-Gershon T, Hsu JY, Kadonaga JT. (2008b). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev*. 22(20):2823-30.

Juven-Gershon T, Kadonaga JT. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol*. 339(2):225-9.

Kadauke S, Blobel GA. (2009). Chromatin loops in gene regulation. *Biochim Biophys Acta*. 1789(1):17-25.

Kadonaga JT. (2002). The DPE, a core promoter element for transcription by RNA polymerase II. *Exp Mol Med*. 34(4):259-64.

Kain SR, Ganguly S. (2001). Overview of genetic reporter systems. *Curr Protoc Mol Biol*. Chapter 9:Unit 9.6.

Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH. (2000). Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum Mol Genet*. 9(17):2563-72.

Kamp C. (2001). Der Spermatogenese-Locus AZFa des humanen Y-Chromosoms: Molekulare Bruchpunkt-Kartierung und ein Beitrag zur funktionellen Charakterisierung des DEAD-Box-Y-Gens. Dissertation, vorgelegt der Fakultät für Biowissenschaften der Ruprecht-Karls-Universität Heidelberg.

Kanai Y, Hiramatsu R, Matoba S, Kidokoro T. (2005). From SRY to SOX9: mammalian testis differentiation. *J Biochem*. 138(1):13-9.

Kankia BI, Barany G, Musier-Forsyth K. (2005). Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic Acids Res*. 33(14):4395-403.

Kaplan J, Calame K. (1997). The ZiN/POZ domain of ZF5 is required for both transcriptional activation and repression. *Nucleic Acids Res*. 25(6):1108-16.

Kapustin Y, Souvorov A, Tatusova T, Lipman D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct*. 3:20.

Kashimada K, Koopman P. (2010). Sry: the master switch in mammalian sex determination. *Development*. 137(23):3921-30.

Kato S. (2000). Molecular mechanism of transcriptional control by nuclear vitamin receptors. *Br J Nutr*. 84 Suppl 2:S229-33.

Katsani KR, Hajibagheri MA, Verrijzer CP. (1999). Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J*. 18(3):698-708.

Kaur P, Kaur G, Bansal MP. (2008). Upregulation of AP1 by tertiary butyl hydroperoxide induced oxidative stress and subsequent effect on spermatogenesis in mice testis. *Mol Cell Biochem*. 308(1-2):177-81.

Kawaji H, Frith MC, Katayama S, Sandelin A, Kai C, Kawai J, Carninci P, Hayashizaki Y. (2006). Dynamic usage of transcription start sites within core promoters. *Genome Biol*. 7(12):R118.

Kawamata M, Inoue H, Nishimori K. (2007). Male-specific function of Dmrt7 by sexually dimorphic translation in mouse testis. *Sex Dev*. 1(5):297-304.

Kay V, Bode J. (1994). Binding specificity of a nuclear scaffold: supercoiled, single-stranded, and scaffold-attached-region DNA. *Biochemistry*. 33(1):367-74.

Kays AR, Schepartz A. (2000). Virtually unidirectional binding of TBP to the AdMLP TATA box within the quaternary complex with TFIIA and TFIIB. *Chem Biol*. 7(8):601-10.

Kehrer-Sawatzki H, Cooper DN. (2007a). Structural divergence between the human and chimpanzee genomes. *Hum Genet.* 120(6):759-78.

Kehrer-Sawatzki H, Cooper DN. (2007b).Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat*. 28(2):99-130.

Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res*. 31(13):3576-9.

Kelly KF, Daniel JM. (2006). POZ for effect - POZ-ZF transcription factors in cancer and development. *Trends Cell Biol*. 16(11):578-87.

Kennedy GC, German MS, Rutter WJ. (1995). The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat Genet*. 9(3):293-8.

Khachane AN, Harrison PM. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics*. 10:435.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*. 453(7191):56-64.

Kikin O, D'Antonio L, Bagga PS. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*. 34(Web Server issue):W676-82.

Kim S, Namekawa SH, Niswander LM, Ward JO, Lee JT, Bardwell VJ, Zarkower D. (2007). A mammal-specific Doublesex homolog associates with male sex chromatin and is required for male meiosis. *PLoS Genet*. 3(4):e62.

Kim SH, Sussman JL, Suddath FL, Quigley GJ, McPherson A, Wang AH, Seeman NC, Rich A. (1974). The general structure of transfer RNA molecules. *Proc Natl Acad Sci USA*. 71(12):4970-4.

Kim TM, Chung YJ, Rhyu MG, Jung MH. (2007). Germline methylation patterns inferred from local nucleotide frequency of repetitive sequences in the human genome. *Mamm Genome*. 18(4):277-85.

Kim YS, Lee SG, Park SH, Song K. (2001). Gene structure of the human DDX3 and chromosome mapping of its related sequences. *Mol Cells*. 12(2):209-14.

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, et al. (2006). Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res*. 16(1):55-65.

Kimura M. (1983). The neutral theory of molecular evolution. *Cambridge University Press*, Cambridge.

Kimura T, Yomogida K, Iwai N, Kato Y, Nakano T. (1999). Molecular cloning and genomic organization of mouse homologue of Drosophila germ cell-less and its expression in germ lineage cells. *Biochem Biophys Res Commun*. 262(1):223-30.

Kingsley C, Winoto A. (1992). Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol Cell Biol*. 12(10):4251-61.

Kinuta K, Tanaka H, Moriwake T, Aya K, Kato S, Seino Y. (2000). Vitamin D is an important factor in estrogen biosynthesis of both female and male gonads. *Endocrinology*. 141(4):1317-24.

Kleene KC. (1989). Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. *Development*. 106(2):367-73.

Kleene KC. (1996). Patterns of translational regulation in the mammalian testis. *Mol Reprod Dev*. 43(2):268-81.

Kleene KC. (2001). A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech Dev*. 106(1-2):3-23.

Kleene KC. (2003). Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenet Genome Res*. 103(3-4):217-24.

Kleene KC, Bagarova J, Hawthorne SK, Catado LM. (2010). Quantitative analysis of mRNA translation in mammalian spermatogenic cells with sucrose and Nycodenz gradients. *Reprod Biol Endocrinol*. 8:155.

Kleinjan DJ, van Heyningen V. (1998). Position effect in human genetic disease. *Hum Mol Genet*. 7: 1611-8.

Kleinjan DA, Seawright A, Schedl A, Quinlan RA, Danes S, van Heyningen V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet*. 10: 2049-59.

Kleinjan DA, van Heyningen V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*. 76(1):8-32.

Klenova EM, Morse HC 3rd, Ohlsson R, Lobanenkov VV. (2002). The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin Cancer Biol*. 12(5):399-414.

Klingenhoff A, Frech K, Quandt K, Werner T. (1999). Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. *Bioinformatics*. 15(3):180-6.

Klostermeier D, Rudolph MG. (2009). A novel dimerization motif in the C-terminal domain of the Thermus thermophilus DEAD box helicase Hera confers substantial flexibility. *Nucleic Acids Res*. 37(2):421-30.

Knippers R. (2001). *Molekulare Genetik*. Georg Thieme Verlag, Stuttgart, 8. Auflage, pp. 334, 392.

Kocer A, Reichmann J, Best D, Adams IR. (2009). Germ cell sex determination in mammals. *Mol Hum Reprod*. 15(4):205-13.

Kochetov AV. (2005). AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. *Bioinformatics*. 21(7):837-40.

Kodzius R, Matsumura Y, Kasukawa T, Shimokawa K, Fukuda S, Shiraki T, Nakamura M, Arakawa T, Sasaki D, Kawai J, Harbers M, Carninci P, Hayashizaki Y. (2004). Absolute expression values for mouse transcripts: re-annotation of the READ expression database by the use of CAGE and EST sequence tags. *FEBS Lett*. 559(1-3):22-6.

Koerner MV, Barlow DP. (2010). Genomic imprinting-an epigenetic gene-regulatory model. *Curr Opin Genet Dev*. 20(2):164-70.

Kohn M, Kehrer-Sawatzki H, Vogel W, Graves JA, Hameister H. (2004). Wide genome comparisons reveal the origins of the human X chromosome. *Trends Genet*. 20(12):598-603.

Kohwi-Shigematsu T, Maass K, Bode J. (1997). A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry*. 36(40):12005-10.

Kondoh H, Kamachi Y. (2010). SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol*. 42(3):391-9.

Kopytova DV, Krasnov AN, Kopantceva MR, Nabirochkina EN, Nikolenko JV, Maksimenko O, Kurshakova MM, Lebedeva LA, Yerokhin MM, Simonova OB, et al. (2006). Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. *Mol Cell Biol*. 26(20):7492-505.

Korneev SA, Straub V, Kemenes I, Korneeva EI, Ott SR, Benjamin PR, O'Shea M. (2005). Timed and targeted differential regulation of nitric oxide synthase (NOS) and anti-NOS genes by reward conditioning leading to long-term memory formation. *J Neurosci*. 25(5):1188-92.

Kostadinov R, Malhotra N, Viotti M, Shine R, D'Antonio L, Bagga P. (2006). GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res*. 34(Database issue):D119-24.

Kotaja N, Sassone-Corsi P. (2007). The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol*. 8(1):85-90.

Kozak M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res*. 12(2):857-72.

Kozak M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 44(2):283-92.

Kozak M. (1987a). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol*. 196(4):947-50.

Kozak M. (1987b). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*. 15(20):8125-48.

Kozak M. (1989). Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol*. 9(11):5134-42.

Kozak M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem*. 266(30):19867-70.

Kozak M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*. 299(1-2):1-34.

Kozak M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*. Nov 21;361:13-37.

Kozlowski P, Sobczak K, Krzyzosiak WJ. (2010). Trinucleotide repeats: triggers for genomic disorders? *Genome Med*. 2(4):29.

Krausz C, Degl'Innocenti S, Nuti F, Morelli A, Felici F, Sansone M, Varriale G, Forti G. (2006). Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. *Hum Mol Genet*. 15(18):2673-81.

Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. (2007). Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet*. 23(4):158-61.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J. (2006). Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol*. 4(4):e91.

Krishnan SP, Liang SS, Veeravalli B. (2010). Towards high performance computing for molecular structure prediction using IBM Cell Broadband Engine - an implementation perspective. *BMC Bioinformatics*. 11 Suppl 1:S36.

Kumari S, Bugaut A, Balasubramanian S. (2008). Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry*. 47(48):12664-9.

Kumari S, Bugaut A, Huppert JL, Balasubramanian S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*. 3(4):218-21.

Kutach AK, Kadonaga JT. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Mol Cell Biol*. 20(13):4754-64.

Kvikstad EM, Makova KD. (2010). The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. *Genome Res*. 20(5):600-13.

Kwong AD, Rao BG, Jeang KT. (2005). Viral and cellular RNA helicases as antiviral targets. *Nat Rev Drug Discov*. 4(10):845-53.

Kypr J, Kejnovská I, Vorlícková M. (2007). Conformations of DNA strands containing GAGT, GACA, or GAGC tetranucleotide repeats. *Biopolymers*. 87(4):218-24.

Laaser I, Theis FJ, de Angelis MH, Kolb HJ, Adamski J. (2011). Huge splicing frequency in human Y chromosomal UTY gene. *OMICS*. 15(3):141-54.

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev*. 12(1):34-44.

Lahn BT, Page DC. (1997). Functional coherence of the human Y chromosome. *Science*. 278(5338):675-80.

Lahn BT, Page DC. (1999). Four evolutionary strata on the human X chromosome. *Science*. 286(5441):964-7.

Lai MC, Lee YH, Tarn WY. (2008). The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. *Mol Biol Cell*. 19(9):3847-58.

Lai MC, Chang WC, Shieh SY, Tarn WY. (2010). DDX3 regulates cell growth through translational control of cyclin E1. *Mol Cell Biol*. 30(22):5444-53.

Lamph WW, Dwarki VJ, Ofir R, Montminy M, Verma IM. (1990). Negative and positive regulation by transcription factor cAMP response element-binding protein is modulated by phosphorylation. *Proc Natl Acad Sci USA*. 87(11):4320-4.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. (2001). Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860-921.

Lang D, Epstein JA. (2003). Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. *Hum Mol Genet*. 12(8):937-45.

Lang D, Powell SK, Plummer RS, Young KP, Ruggeri BA. (2007). PAX genes: roles in development, pathophysiology, and cancer. *Biochem Pharmacol*. 73(1):1-14.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. 23(21):2947-8.

Lau YFC, Kido T, Li Y. (2007). The *TSPY* gene family. In Lau YFC, Chan WY (ed.), *The Y Chromosome and Male Germ Cell Biology in Health and Diseases*. World Scientific Publishing Company, Singapur, Singapur, 1. Auflage, pp. 73-90.

Lawson MJ, Zhang L. (2008). Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene*. 407(1-2):54-62.

Leatherman JL, Kaestner KH, Jongens TA. (2000). Identification of a mouse germ cell-less homologue with conserved activity in Drosophila. *Mech Dev*. 92(2):145-53.

Lee CS, Dias AP, Jedrychowski M, Patel AH, Hsu JL, Reed R. (2008). Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res*. 36(14):4708-18.

Lee JS, Galvin KM, Shi Y. (1993). Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc Natl Acad Sci USA*. 90(13):6145-9.

Lee JY, Ji Z, Tian B. (2008). Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res*. 36(17):5581-90.

Lee KH, Lee YS. (2004). Cloning and characterization of zinc finger protein 161 in Rattus norvegicus. *DNA Seq*. 15(4):310-3.

Lee LT, Lam IP, Chow BK. (2008). A functional variable number of tandem repeats is located at the 5' flanking region of the human secretin gene plays a downregulatory role in expression. *J Mol Neurosci*. 36(1-3):125-31.

Lehmann M. (2004). Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet*. 20(1):15-22.

Lemos B. (2007). The Opossum genome reveals further evidence for regulatory evolution in mammalian diversification. *Genome Biol*. 8(8):223.

Lenartowski R, Goc A. (2002). Tissue-specific association of the human tyrosine hydroxylase gene with the nuclear matrix. *Neurosci Lett*. 330(2):151-4.

Leroy P, Alzari P, Sassoon D, Wolgemuth D, Fellous M. (1989). The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. Cell. 57(4):549-59.

Leroy P, Seboun E, Mattei MG, Fellous M, Bishop CE. (1987). Testis-specific transcripts detected by a human Y-DNA-derived probe. *Development*. 101 Suppl:177-83.

Lettice LA, Hill RE. (2005). Preaxial polydactyly: a model for defective long-range regulation in congenital abnormalities. *Curr Opin Genet Dev*. 15(3):294-300.

Lévesque L, Bor YC, Matzat LH, Jin L, Berberoglu S, Rekosh D, Hammarskjöld ML, Paschal BM. (2006). Mutations in tap uncouple RNA export activity from translocation through the nuclear pore complex. *Mol Biol Cell*. 17(2):931-43.

Levine M, Tjian R. (2003). Transcription regulation and animal diversity. *Nature*. 424: 147-51.

Li B, Carey M, Workman JL. (2007). The role of chromatin during transcription. *Cell*. 128(4):707-19.

Li H, Liu H, Wang Z, Liu X, Guo L, Huang L, Gao L, McNutt MA, Li G. (2008). The role of transcription factors Sp1 and YY1 in proximal promoter region in initiation of transcription of the mu opioid receptor gene in human lymphocytes. *J Cell Biochem*. 104(1):237-50.

Li J, Hawkins IC, Harvey CD, Jennings JL, Link AJ, Patton JG. (2003). Regulation of alternative splicing by SRrp86 and its interacting proteins. *Mol Cell Biol*. 23(21):7437-47.

Li Q, Gardner DG. (1994). Negative regulation of the human atrial natriuretic peptide gene by 1,25-dihydroxyvitamin D3. *J Biol Chem*. 269(7):4934-9.

Li WH. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol*. 36(1):96-9.

Li WH. (1997). *Molecular Evolution*. Sinauer Associates, Sunderland, USA.

Lian Z, Karpikov A, Lian J, Mahajan MC, Hartman S, Gerstein M, Snyder M, Weissman SM. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome Res*. 18(8):1224-37.

Liang H, Lin YS, Li WH. (2008). Fast evolution of core promoters in primate genomes. *Mol Biol Evol*. 25(6):1239-44.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 326(5950):289-93.

Liebich I, Bode J, Frisch M, Wingender E. (2002). S/MARt DB: a database on scaffold/matrix attached regions. *Nucleic Acids Res*. 30(1):372-4.

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev*. 18(13):1606-17.

Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN, Phan AT. (2009). Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc*. 131(12):4301-9.

Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K, Schnier J, Slonimski PP. (1989). Birth of the D-E-A-D box. *Nature*. 337(6203):121-2.

Linder P. (2006). Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res*. 34(15):4168-80.

Lindqvist A. (2010). Cyclin B-Cdk1 activates its own pump to get into the nucleus. *J Cell Biol*. 189(2):197-9.

Linnemann AK, Platts AE, Krawetz SA. (2009). Differential nuclear scaffold/matrix attachment marks expressed genes. *Hum Mol Genet*. 18(4):645-54.

Litt M, Luty JA. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet*. 44(3):397-401.

Liu F, Putnam A, Jankowsky E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A*. 105(51):20209-14.

Liu H, Deng X, Shyu YJ, Li JJ, Taparowsky EJ, Hu CD. (2006). Mutual regulation of c-Jun and ATF2 by transcriptional activation and subcellular localization. *EMBO J*. 25(5):1058-69.

Liu H, Shi Y. (2005).Yin Yang 1. In Iuchi S, Kuldell N. (ed.), *Zinc Finger Proteins. From Atomic Contact to Cellular Function.* Landes Bioscience/Eurekah.com und Kluwer Academic/Plenum Publishers, New York, USA, pp. 182-194.

Liu H, Schmidt-Supprian M, Shi Y, Hobeika E, Barteneva N, Jumaa H, Pelanda R, Reth M, Skok J, Rajewsky K, Shi Y. (2007). Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev*. 21(10):1179-89.

Liu WS, Wang A, Yang Y, Chang TC, Landrito E, Yasue H. (2009). Molecular characterization of the DDX3Y gene and its homologs in cattle. *Cytogenet Genome Res*. 126(4):318-28.

Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*. 288(5463):136-40.

Loots GG. (2008). Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Adv Genet*. 61:269-93.

Lopes AM, Miguel RN, Sargent CA, Ellis PJ, Amorim A, Affara NA. (2010). The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis. *BMC Mol Biol*. 11:33.

Luddi A, Margollicci M, Gambera L, Serafini F, Cioni M, De Leo V, Balestri P, Piomboni P. (2009). Spermatogenesis in a man with complete deletion of USP9Y. *N Engl J Med*. 360(9):881-5.

Ludérus ME, den Blaauwen JL, de Smit OJ, Compton DA, van Driel R. (1994). Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. *Mol Cell Biol*. 14(9):6297-305.

Luetjens CM, Weinbauer GF, Wistuba J. (2005). Primate spermatogenesis: new insights into comparative testicular organisation, spermatogenic efficiency and endocrine control. *Biol Rev Camb Philos Soc*. 80(3):475-88.

Lutz CS. (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol*. 3(10):609-17.

Lutz CS, Moreira A. (2011). Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *WIREs RNA*. 2(1):23-31.

Luukko K, Ylikorkala A, Mäkelä TP. (2001). Developmentally regulated expression of Smad3, Smad4, Smad6, and Smad7 involved in TGF-beta signaling. *Mech Dev*. 101(1-2):209-12.

Lysiak JJ, Kirby JL, Tremblay JJ, Woodson RI, Reardon MA, Palmer LA, Turner TT. (2009). Hypoxia-inducible factor-1alpha is constitutively expressed in murine Leydig cells and regulates 3beta-hydroxysteroid dehydrogenase type 1 promoter activity. *J Androl*. 30(2):146-56.

Ma X, Li-Ling J, Huang Q, Chen X, Hou L, Ma F. (2009). Systematic analysis of alternative promoters correlated with alternative splicing in human genes. *Genomics*. 93(5):420-5.

Malik S, Roeder RG. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet*. 11(11):761-72.

Manna PR, Dyson MT, Eubank DW, Clark BJ, Lalli E, Sassone-Corsi P, Zeleznik AJ, Stocco DM. (2002). Regulation of steroidogenesis and the steroidogenic acute regulatory protein by a member of the cAMP response-element binding protein family. *Mol Endocrinol*. 16(1):184-99.

Mantovani R. (1998). A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res*. 26(5):1135-43.

Margalit A, Liu J, Fridkin A, Wilson KL, Gruenbaum Y. (2005). A lamin-dependent pathway that regulates nuclear organization, cell cycle progression and germ cell development. *Novartis Found Symp*. 264:231-40; discussion 240-5.

Marshall Graves JA. (2008). Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev Genet*. 42:565-86.

Martin RD, Soligo C, Tavaré S. (2007). Primate origins: implications of a cretaceous ancestry. *Folia Primatol (Basel)*. 78(5-6):277-96.

Martin-Coello J, Dopazo H, Arbiza L, Ausió J, Roldan ER, Gomendio M. (2009). Sexual selection drives weak positive selection in protamine genes and high promoter divergence, enhancing sperm competitiveness. *Proc Biol Sci*. 276(1666):2427-36.

Martinez E. (2010). The NF-Y angle to the CCAAT's tale. *Cell Cycle*. 9(4):642-3.

Martinez-Morales PL, Quiroga AC, Barbas JA, Morales AV. (2010). SOX5 controls cell cycle progression in neural progenitors by interfering with the WNT-beta-catenin pathway. *EMBO Rep*. 11(6):466-72.

Martins WS, Lucas DC, Neves KF, Bertioli DJ. (2009). WebSat - a web software for microsatellite marker development. *Bioinformation*. 3(6):282-3.

Maruyama K, Sugano S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*. 138(1-2):171-4.

Mašek T, Valášek L, Pospíšek M. (2011). Polysome analysis and RNA purification from sucrose gradients. *Methods Mol Biol*. 703:293-309.

Mata J, Marguerat S, Bähler J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci*. 30(9):506-14.

Matharu NK, Hussain T, Sankaranarayanan R, Mishra RK. (2010). Vertebrate homologue of Drosophila GAGA factor. *J Mol Biol*. 400(3):434-47.

Mathews DH, Sabina J, Zuker M & Turner DH. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure *J. Mol. Biol.* 288(5):911-940.

Matis C, Chomez P, Picard J, Rezsohazy R. (2001). Differential and opposed transcriptional effects of protein fusions containing the VP16 activation domain. *FEBS Lett*. 499(1-2):92-6.

Matsuda M, Shinomiya A, Kinoshita M, Suzuki A, Kobayashi T, Paul-Prasanth B, Lau EL, Hamaguchi S, Sakaizumi M, Nagahama Y. (2007). DMY gene induces male development in genetically female (XX) medaka fish. *Proc Natl Acad Sci USA*. 104(10):3865-70.

Matsui A, Rakotondraparany F, Munechika I, Hasegawa M, Horai S. (2009). Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. *Gene*. 441(1-2):53-66.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 34(Database issue):D108-10.

Mayr E. (2001). *Das ist Evolution*. 3. Auflage 2005. Wilhelm Goldmann Verlag, München.

Mazeyrat S, Saut N, Sargent CA, Grimmond S, Longepied G, Ehrmann IE, Ellis PS, Greenfield A, Affara NA, Mitchell MJ. (1998). The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. *Hum Mol Genet*. 7(11):1713-24.

McKeon C, Schmidt A, de Crombrugghe B. (1984). A sequence conserved in both the chicken and mouse alpha 2(I) collagen promoter contains sites sensitive to S1 nuclease. *J Biol Chem*. 259(10):6636-40.

McKnight RA, Shamay A, Sankaran L, Wall RJ, Hennighausen L. (1992). Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proc Natl Acad Sci USA*. 89(15):6943-7.

McLuckie KI, Waller ZA, Sanders DA, Alves D, Rodriguez R, Dash J, McKenzie GJ, Venkitaraman AR, Balasubramanian S. (2011). G-quadruplex-binding benzo[a]phenoxazines down-regulate c-KIT expression in human gastric carcinoma cells. *J Am Chem Soc*. 133(8):2658-63.

McMurray CT. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*. 11(11):786-99.

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, et al. (2001). A physical map of the human genome. *Nature*. 409(6822):934-41.

McVean G. (2010). What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc Lond B Biol Sci*. 365(1544):1213-8.

Meerovitch K, Nicholson R, Sonenberg N. (1991). In vitro mutational analysis of cis-acting RNA translational elements within the poliovirus type 2 5' untranslated region. *J Virol*. 65(11):5895-901.

Melo EO, de Melo Neto OP, Martins de Sá C. (2003). Adenosine-rich elements present in the 5'-untranslated region of PABP mRNA can selectively reduce the abundance and translation of CAT mRNAs in vivo. *FEBS Lett*. 546(2-3):329-34.

Merika M, Thanos D. (2001). Enhanceosomes. *Curr Opin Genet Dev*. 11(2):205-8.

Mertin S, McDowall SG, Harley VR. (1999). The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res*. 27(5):1359-64.

Merz C, Urlaub H, Will CL, Lührmann R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA*. 13(1):116-28.

Metzger KJ. (2011). Helping Students Conceptualize Species Divergence Events using the Online Tool “TimeTree: The Timescale of Life”. *The American Biology Teacher*. 73(2):106-108.

Meyuhas O. (2000). Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem*. 267(21):6321-30.

Mezquita P, Mezquita C, Mezquita J. (1999). Novel transcripts of carbonic anhydrase II in mouse and human testis. *Mol Hum Reprod*. 5(3):199-205.

Michelhaugh SK, Fiskerstrand C, Lovejoy E, Bannon MJ, Quinn JP. (2001). The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons. *J Neurochem*. 79(5):1033-8.

Mielke C, Christensen MO, Westergaard O, Bode J, Benham CJ, Breindl M. (2002). Multiple collagen I gene regulatory elements have sites of stress-induced DNA duplex destabilization and nuclear scaffold/matrix association potential. *J Cell Biochem*. 84(3):484-96.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, et al. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature*. 447(7141):167-77.

Millar MR, Sharpe RM, Weinbauer GF, Fraser HM, Saunders PT. (2000). Marmoset spermatogenesis: organizational similarities to the human. *Int J Androl*. 23(5):266-77.

Mirkin SM. (2007). Expandable DNA repeats and human disease. *Nature*. 447(7147):932-40.

Miyahara K, Kuge H, Shizuta Y, Honke K. (2004). Three repeats of CCCCTCC on the pyrimidine-rich sequence in the proximal 5' flanking region are required for efficient transcriptional activity of the human endothelial nitric oxide synthase gene. *Free Radic Res*. 38(1):87-95.

Miyazaki T, Arai S. (2007). Two distinct controls of mitotic cdk1/cyclin B1 activity requisite for cell growth prior to cell division. *Cell Cycle*. 6(12):1419-25.

Mochizuki K, Nishimiya-Fujisawa C, Fujisawa T. (2001). Universal occurrence of the vasa-related genes among metazoans and their germline expression in Hydra. *Dev Genes Evol*. 211(6):299-308.

Monaco C, Helmer Citterich M, Caprini E, Vorechovsky I, Russo G, Croce CM, Barbanti-Brodano G, Negrini M. (1998). Molecular cloning and characterization of ZNF202: a new gene at 11q23.3 encoding testis-specific zinc finger proteins. *Genomics*. 52(3):358-62.

Morachis JM, Murawsky CM, Emerson BM. (2010). Regulation of the p53 transcriptional response by structurally diverse core promoters. Genes Dev. 24(2):135-47.

Morgenstern B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*. 15(3):211-8.

Morley SJ, Coldwell MJ. (2008). A cunning stunt: an alternative mechanism of eukaryotic translation initiation. *Sci Signal*. 1(25):pe32.

Möröy T, Geisen C. (2004). Cyclin E. *Int J Biochem Cell Biol*. 36(8):1424-39.

Morris MJ, Negishi Y, Pazsint C, Schonhoft JD, Basu S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J Am Chem Soc*. 132(50):17831-9.

Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JM. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet*. 40(6):794-9.

Muir T, Wilson-Rawls J, Stevens JD, Rawls A, Schweitzer R, Kang C, Skinner MK. (2008). Integration of CREB and bHLH transcriptional signaling pathways through direct heterodimerization of the proteins: role in muscle and testis development. *Mol Reprod Dev*. 75(11):1637-52.

Müller F, Zaucker A, Tora L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. *Curr Opin Genet Dev*. 20(5):533-40.

Müller GA, Engeland K. (2010). The central role of CDE/CHR promoter elements in the regulation of cell cycle-dependent gene transcription. *FEBS J*. 277(4):877-93.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51: 263-73.

Munk K. (2010). *Taschenlehrbuch Biologie: Genetik*. Georg Thieme Verlag, Stuttgart, 1. Auflage, pp. 14-15.

Murayama A, Takeyama K, Kitanaka S, Kodera Y, Kawaguchi Y, Hosoya T, Kato S. (1999). Positive and negative regulations of the renal 25-hydroxyvitamin D3 1alpha-hydroxylase gene by parathyroid hormone, calcitonin, and 1alpha,25(OH)2D3 in intact animals. *Endocrinology*. 140(5):2224-31.

Murphy MW, Zarkower D, Bardwell VJ. (2007). Vertebrate DM domain proteins bind similar DNA sequences and can heterodimerize on DNA. *BMC Mol Biol*. 8:58.

Murphy WJ, Sun S, Chen ZQ, Pecon-Slattery J, O'Brien SJ. (1999). Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping.*Genome Res*. 9(12):1223-30.

Nakagomi K, Kohwi Y, Dickinson LA, Kohwi-Shigematsu T. (1994). A novel DNA-binding motif in the nuclear matrix attachment DNA-binding protein SATB1. *Mol Cell Biol*. 14(3):1852-60.

Nakamura Y, Koyama K, Matsushima M. (1998). VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. *J Hum Genet*. 43(3):149-52.

Nam J, Dong P, Tarpine R, Istrail S, Davidson EH. (2010). Functional cis-regulatory genomics for systems biology. *Proc Natl Acad Sci USA*. 107(8):3930-5.

Namciu SJ, Blochlinger KB, Fournier RE. (1998). Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. *Mol Cell Biol*. 18(4):2382-91.

Namekawa SH, Park PJ, Zhang LF, Shima JE, McCarrey JR, Griswold MD, Lee JT. (2006). Postmeiotic sex chromatin in the male germline of mice. *Curr Biol*. 16(7):660-7.

Nanassy L, Griffin J, Emery BR, Carrell DT. (2010). The marmoset and cotton rat as animal models for the study of sperm chromatin packaging. *Syst Biol Reprod Med*. 56(3):207-12.

Navarro-Costa P, Plancha CE, Gonçalves J. (2010). Genetic dissection of the AZF regions of the human Y chromosome: thriller or filler for male (in)fertility? *J Biomed Biotechnol*. 2010:936569.

Nayler O, Strätling W, Bourquin JP, Stagljar I, Lindemann L, Jasper H, Hartmann AM, Fackelmayer FO, Ullrich A, Stamm S. (1998). SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res*. 26(15):3542-9.

Neafsey DE, Galagan JE. (2007). Dual modes of natural selection on upstream open reading frames. *Mol Biol Evol*. 24(8):1744-51.

Neidle S. (2009). The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struct Biol*. 19(3):239-50.

Neil DL, Villasante A, Fisher RB, Vetrie D, Cox B, Tyler-Smith C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res*. 18(6):1421-8.

Newnham CM, Hall-Pogar T, Liang S, Wu J, Tian B, Hu J, Lutz CS. (2010). Alternative polyadenylation of MeCP2: Influence of cis-acting elements and trans-acting factors. *RNA Biol*. 7(3):361-72.

Nielsen SJ, Praestegaard M, Jorgensen HF, Clark BF. (1998). Different Sp1 family members differentially affect transcription from the human elongation factor 1 A-1 gene promoter. *Biochem J*. 333 (Pt 3):511-7.

Ning Z, Cox AJ, Mullikin JC. (2001). SSAHA: a fast search method for large DNA databases. *Genome Res*. 11(10):1725-9.

Noce T, Okamoto-Ito S, Tsunekawa N. (2001). Vasa homolog genes in mammalian germ cell development. *Cell Struct Funct*. 26(3):131-6.

Nonaka M. (2009). A Janus-like role of CREB protein: enhancement of synaptic property in mature neurons and suppression of synaptogenesis and reduced network synchrony in early development. *J Neurosci*. 29(20):6389-91.

Norman AW. (2006). Minireview: vitamin D receptor: new assignments for an already busy receptor. *Endocrinology*. 147(12):5542-8.

Numoto M, Yokoro K, Koshi J. (1999). ZF5, which is a Kruppel-type transcriptional repressor, requires the zinc finger domain for self-association. *Biochem Biophys Res Commun*. 256(3):573-8.

Obata T, Yanagidani A, Yokoro K, Numoto M, Yamamoto S. (1999). Analysis of the consensus binding sequence and the DNA-binding domain of ZF5. *Biochem Biophys Res Commun*. 255(2):528-34.

Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E. (2007*).* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet*. 39(6):730-2.

Ogilvie VC, Wilson BJ, Nicol SM, Morrice NA, Saunders LR, Barber GN, Fuller-Pace FV. (2003). The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res*. 31(5):1470-80.

Ohler U, Wassarman DA. (2010). Promoting developmental transcription. *Development*. 137(1):15-26.

Ohlsson R, Lobanenkov V, Klenova E. (2010). Does CTCF mediate between nuclear organization and gene expression? *Bioessays*. 32(1):37-50.

Olive V, Wagner N, Chan S, Kastner P, Vannetti C, Cuzin F, Rassoulzadegan M. (2007). PU.1 (Sfpi1), a pleiotropic regulator expressed from the first embryonic stages with a crucial function in germinal progenitors. *Development*. 134(21):3815-25.

Oliver AW, Bogdarina I, Schroeder E, Taylor IA, Kneale GG. (2000). Preferential binding of fd gene 5 protein to tetraplex nucleic acid structures. *J Mol Biol*. 301(3):575-84.

Onohara Y, Fujiwara T, Yasukochi T, Himeno M, Yokota S. (2010). Localization of mouse vasa homolog protein in chromatoid body and related nuage structures of mammalian spermatogenic cells during spermatogenesis. *Histochem Cell Biol*. 133(6):627-39.

Oostra BA, Willemsen R. (2009). FMR1: a gene with three faces. *Biochim Biophys Acta*. 1790(6):467-77.

Orlov SV, Kuteykin-Teplyakov KB, Ignatovich IA, Dizhe EB, Mirgorodskaya OA, Grishin AV, Guzhova OB, Prokhortchouk EB, Guliy PV, Perevozchikov AP. (2007). Novel repressor of the human FMR1 gene - identification of p56 human (GCC)(n)-binding protein as a Krüppel-like transcription factor ZF5. *FEBS J*. 274(18):4848-62.

Orr HT, Zoghbi HY. (2007). Trinucleotide repeat disorders. *Annu Rev Neurosci*. 30:575-621.

Osaki E, Nishina Y, Inazawa J, Copeland NG, Gilbert DJ, Jenkins NA, Ohsugi M, Tezuka T, Yoshida M, Semba K. (1999). Identification of a novel Sry-related gene and its germ cell-specific expression. *Nucleic Acids Res*. 27(12):2503-10.

Ottaviani D, Lever E, Takousis P, Sheer D. (2008). Anchoring the genome. *Genome Biol*. 9(1):201.

Õunap K, Uibo O, Zordania R, Kiho L, Ilus T, Oiglane-Shlik E, Bartsch O. (2004). Three patients with 9p deletions including DMRT1 and DMRT2: a girl with XY complement, bilateral ovotestes, and extreme growth retardation, and two XX females with normal pubertal development. *Am J Med Genet A*. 130A(4):415-23.

Oyama M, Itagaki C, Hata H, Suzuki Y, Izumi T, Natsume T, Isobe T, Sugano S. (2004). Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. *Genome Res*. 14(10B):2048-52.

Oyama M, Kozuka-Hata H, Suzuki Y, Semba K, Yamamoto T, Sugano S. (2007). Diversity of translation start sites may define increased complexity of the human short ORFeome. *Mol Cell Proteomics*. 6(6):1000-6.

Pääbo, S. (2010). Vergleichende Genomik. In Forschungsperspektiven 2010+ der Max-Planck-Gesellschaft, pp. 36 - 37.

Paguio A, Almond B, Fan F, Stecha P, Garvin D, Wood M, Wood K. (2005). pGL4 Vectors: A New Generation of Luciferase Reporter Vectors*. Promega Notes* 89:7–10.

Palladino MA, Pirlamarla PR, McNamara J, Sottas CM, Korah N, Hardy MP, Hales DB, Hermo L. (2011). Normoxic expression of hypoxia-inducible factor 1 in rat Leydig cells in vivo and in vitro. *J Androl*. 32(3):307-23.

Pan C, Russell R. (2010). Roles of DEAD-box proteins in RNA and RNP Folding. *RNA Biol*. 7(6):667-76.

Pan Y, Tsai CJ, Ma B, Nussinov R. (2010). Mechanisms of transcription factor selectivity. *Trends Genet*. 26(2):75-83.

Panne D. (2008). The enhanceosome. *Curr Opin Struct Biol*. 18(2):236-42.

Papapetrou EP, Ziros PG, Micheva ID, Zoumbos NC, Athanassiadou A. (2006). Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. *Gene Ther*. 13(1):40-51.

Park PJ, Butte AJ, Kohane IS. (2002). Comparing expression profiles of genes with similar promoter regions. *Bioinformatics*. 18(12):1576-84.

Park SH, Lee SG, Kim Y, Song K. (1998). Assignment of a human putative RNA helicase gene, DDX3, to human X chromosome bands p11.3-->p11.23. *Cytogenet Cell Genet*. 81(3-4):178-9.

Parkinson GN. (2006). *Fundamentals of quadruplex structures*. In Neidle S, Balasubramanian S. (ed.), Quadruplex nucleic acids. RSC Publishing, Cambridge, UK, pp.1-30.

Parvinen M. (2005). The chromatoid body in spermatogenesis. *Int J Androl*. 28(4):189-201.

Passarge E, Kohlhase J. (2006). Genetik. In Siegenthaler W, Blum HE (ed.), *Klinische Pathophysiologie*. Georg Thieme Verlag, Stuttgart, 9. Auflage, pp. 4 - 25.

Paul C, Teng S, Saunders PT. (2009). A single, mild, transient scrotal heat stress causes hypoxia and oxidative stress in mouse testes, which induces germ cell death. *Biol Reprod*. 80(5):913-9.

Paulson JR, Laemmli UK. (1977). The structure of histone-depleted metaphase chromosomes. *Cell*. 12(3):817-28.

Pearks Wilkerson AJ, Raudsepp T, Graves T, Albracht D, Warren W, Chowdhary BP, Skow LC, Murphy WJ. (2008). Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. *Genomics*. 92(5):329-38.

Pelletier J, Sonenberg N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell*. 40(3):515-26.

Perera EM, Martin H, Seeherunvong T, Kos L, Hughes IA, Hawkins JR, Berkovitz GD. (2001). Tescalcin, a novel gene encoding a putative EF-hand Ca(2+)-binding protein, Col9a3, and renin are expressed in the mouse testis during the early stages of gonadal differentiation. *Endocrinology*. 142(1):455-63.

Perera EM, Bao Y, Kos L, Berkovitz G. (2010). Structural and functional characterization of the mouse tescalcin promoter. *Gene*. 464(1-2):50-62.

Pérez-Sánchez C, Gómez-Ferrería MA, de La Fuente CA, Granadino B, Velasco G, Esteban-Gamboa A, Rey-Campos J. (2000a). FHX, a novel fork head factor with a dual DNA binding specificity. *J Biol Chem*. 275(17):12909-16.

Pérez-Sánchez C, Arias-de-la-Fuente C, Gómez-Ferrería MA, Granadino B, Rey-Campos J. (2000b). FHX.L and FHX.S, two isoforms of the human fork-head factor FHX (FOXJ2) with differential activity. *J Mol Biol*. 301(4):795-806.

Persengiev SP, Zhu X, Dixit BL, Maston GA, Kittler EL, Green MR. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc Natl Acad Sci USA*. 100(25):14887-91.

Pessler F, Hernandez N. (2003). Flexible DNA binding of the BTB/POZ-domain protein FBI-1. *J Biol Chem*. 278(31):29327-35.

Phillips K, Luisi B. (2000). The virtuoso of versatility: POU proteins that flex to fit. *J Mol Biol*. 302(5):1023-39.

Pickering BM, Willis AE. (2005). The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol*. 16(1):39-47.

Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CU, Pestova TV. (2006). Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev*. 20(5):624-36.

Platts AE, Quayle AK, Krawetz SA. (2006). In-silico prediction and observations of nuclear matrix attachment. *Cell Mol Biol Lett*. 11(2):191-213.

Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban RJ, Hoffmann B, Morkel M, Birchmeier C, Hagemeier C. (2005). A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr Biol*. 15(11):1051-7.

Pointud JC, Mengus G, Brancorsini S, Monaco L, Parvinen M, Sassone-Corsi P, Davidson I. (2003). The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID subunit TAF7, is developmentally regulated during male germ-cell differentiation. *J Cell Sci*. 116(Pt 9):1847-58.

Polavarapu N, Mariño-Ramírez L, Landsman D, McDonald JF, Jordan IK. (2008). Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics*. 9:226.

Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol*. 7(8):R78.

Porter LA, Donoghue DJ. (2003). Cyclin B1 and CDK1: nuclear localization and upstream regulators. *Prog Cell Cycle Res*. 5:335-47.

Prabhakar S, Noonan JP, Pääbo S, Rubin EM. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science*. 314(5800):786.

Prior HM, Walter MA. (1996). SOX genes: architects of development. *Mol Med*. 2(4):405-12.

Quandt K, Frech K, Karas H, Wingender E, Werner T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*. 23(23):4878-84.

Quilter CR, Blott SC, Mileham AJ, Affara NA, Sargent CA, Griffin DK. (2002). A mapping and evolutionary study of porcine sex chromosome genes. *Mamm Genome*. 13(10):588-94.

Qvarnström A, Bailey RI. (2009). Speciation through evolution of sex-linked genes. *Heredity*. 102(1):4-15.

Raab JR, Kamakaka RT. (2010). Insulators and promoters: closer than we think. *Nat Rev Genet*. 11(6):439-46.

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR. (2005). Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol*. 48(3):237-57.

Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, Ohler U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet*. 7(1):e1001274.

Radman-Livaja M, Rando OJ. (2010). Nucleosome positioning: how is it established, and why does it matter? *Dev Biol*. 339(2):258-66.

Rajkowitsch L, Vilela C, Berthelot K, Ramirez CV, McCarthy JE. (2004). Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J Mol Biol*. 335(1):71-85.

Rauschendorf MA. (2006). Identifizierung und Charakterisierung potentieller regulativer Elemente in der 3´-Region des humanen *TWIST1*-Genlocus. Diplomarbeit, vorgelegt dem Fachbereich Biologie der Philipps-Universität Marburg.

Rauschendorf MA, Zimmer J, Hanstein R, Dickemann C, Vogt PH. (2011). Complex transcriptional control of the AZFa gene *DDX3Y* in human testis. *Int J Androl*. 34(1):84-96.

Reeves R. (2000). Structure and function of the HMGI(Y) family of architectural transcription factors. *Environ Health Perspect*. 108 Suppl 5:803-9.

Reina JH, Hernandez N. (2007). On a roll for new TRF targets. *Genes Dev*. 21(22):2855-60.

Reményi A, Schöler HR, Wilmanns M. (2004). Combinatorial control of gene expression. *Nat Struct Mol Biol*. 11(9):812-5.

Renz A, Fackelmayer FO. (1996). Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA. *Nucleic Acids Res*. 24(5):843-9.

Resch AM, Ogurtsov AY, Rogozin IB, Shabalina SA, Koonin EV. (2009). Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics*. 10:162.

Rice P, Longden I, Bleasby A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 16(6):276-7.

Rich A, Nordheim A, Wang AH. (1984). The chemistry and biology of left-handed Z-DNA. *Annu Rev Biochem*. 53:791-846.

Richards JE, Hawley RS. (2010). *The Human Genome: A User's Guide*. Academic Press, Waltham, USA, 3. Auflage.

Richter JD, Sonenberg N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*. 433(7025):477-80.

Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, Fitzgerald P, Vinson C. (2010). CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc Natl Acad Sci USA*. 107(47):20311-6.

Riss J, Laskov R. (1999). Expression of novel alternatively spliced isoforms of the oct-1 transcription factor. *Biochim Biophys Acta*. 1444(2):295-8.

Ritter DI, Li Q, Kostka D, Pollard KS, Guo S, Chuang JH. (2010). The importance of being cis: evolution of orthologous fish and mammalian enhancer activity. *Mol Biol Evol*. 27(10):2322-32.

Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BFC, Klug A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*. 250:546-551.

Rocak S, Linder P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol*. 5(3):232-41.

Rödelsperger C, Köhler S, Schulz MH, Manke T, Bauer S, Robinson PN. (2009). Short ultraconserved promoter regions delineate a class of preferentially expressed alternatively spliced transcripts. *Genomics*. 94(5):308-16.

Rosner A, Rinkevich B. (2007). The DDX3 subfamily of the DEAD box helicases: divergent roles as unveiled by studying different organisms and in vitro assays. *Curr Med Chem*. 14(23):2517-25.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, et al. (2005). The DNA sequence of the human X chromosome. *Nature*. 434(7031):325-37.

Rowold DJ, Herrera RJ. (2000). Alu elements and the human genome. *Genetica*. 108(1):57-72.

Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, FitzGerald PC, Vinson C. (2008). All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. *BMC Genomics*. 9:67.

Rubidge BS, Sidor CA. (2001). Evolutionary patterns among permo-triassic Therapsids. *Annu. Rev. Ecol. Syst*. 32:449–80.

Rubin CM, VandeVoort CA, Teplitz RL, Schmid CW. (1994). Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res*. 22(23):5121-7.

Rudd S, Frisch M, Grote K, Meyers BC, Mayer K, Werner T. (2004). Genome-wide in silico mapping of scaffold/matrix attachment regions in Arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. *Plant Physiol*. 135(2):715-22.

Saffer JD, Jackson SP, Annarella MB. (1991). Developmental expression of Sp1 in the mouse. *Mol Cell Biol*. 11(4):2189-99.

Saga Y, Hata N, Kobayashi S, Magnuson T, Seldin MF, Taketo MM. (1996). MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development*. 122(9):2769-78.

Sambrook, J.Russel, D. (2001). Molecular Cloning. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*. 320(5883):1643-7.

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. (2007). *Nat Rev Genet*. 8(6):424-36.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-7.

Sasaki M, Lange J, Keeney S. (2010). Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol*. 11(3):182-95.

Sassone-Corsi P. (1998). Regulating the balance between differentiation and apoptosis: role of CREM in the male germ cells. *J Mol Med*. 76(12):811-7.

Saunders PT, Gaughan J, Saxty BA, Kerr LE, Millar MR. (1996). Expression of protamine P2 in the testis of the common marmoset and man visualized using non-radioactive in-situ hybridization. *Int J Androl*. 19(4):212-9.

Savinkova LK, Ponomarenko MP, Ponomarenko PM, Drachkova IA, Lysova MV, Arshinova TV, Kolchanov NA. (2009). TATA box polymorphisms in human gene promoters and associated hereditary pathologies. *Biochemistry (Mosc)*. 74(2):117-29.

Sawicka K, Bushell M, Spriggs KA, Willis AE. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans*. 36(Pt 4):641-7.

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, et al. (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 39(Database issue):D38-51.

Scaria V, Hariharan M, Arora A, Maiti S. (2006). Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences. *Nucleic Acids Res*. 34(Web Server issue):W683-5.

Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, et al. (2004). Quality assessment of the human genome sequence. *Nature*. 429(6990):365-8.

Schöler HR, Hatzopoulos AK, Balling R, Suzuki N, Gruss P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor *EMBO J*. 8(9):2543-50.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 132(5):887-98.

Schonhoft JD, Bajracharya R, Dhakal S, Yu Z, Mao H, Basu S. (2009).Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. *Nucleic Acids Res*. 37(10):3310-20.

Schonhoft JD, Das A, Achamyeleh F, Samdani S, Sewell A, Mao H, Basu S. (2010). ILPR repeats adopt diverse G-quadruplex conformations that determine insulin binding. *Biopolymers*. 93(1):21-31.

Schröder M. (2010). Human DEAD-box protein 3 has multiple functions in gene regulation and cell cycle control and is a prime target for viral manipulation. *Biochem Pharmacol*. 79(3):297-306.

Schübeler D, Mielke C, Maass K, Bode J. (1996). Scaffold/matrix-attached regions act upon transcription in a context-dependent manner. *Biochemistry*. 35(34):11160-9.

Schütz W, Alsheimer M, Ollinger R, Benavente R. (2005). Nuclear envelope remodeling during mouse spermiogenesis: postmeiotic expression and redistribution of germline lamin B3. *Exp Cell Res*. 307(2):285-91.

Schütze T, Rubelt F, Repkow J, Greiner N, Erdmann VA, Lehrach H, Konthur Z, Glökler J. (2011). A streamlined protocol for emulsion polymerase chain reaction and subsequent purification. Anal Biochem. 410(1):155-7.

Schwartz S, Ast G. (2010). Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J*. 29(10):1629-36.

Schwartz YB, Pirrotta V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet*. 8(1):9-22.

Segal E, Widom J. (2009a). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol*. 19(1):65-71.

Segal E, Widom J. (2009b). What controls nucleosome positions? *Trends Genet*. 25(8):335-43.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. (2008). Divergent transcription from active promoters. *Science*. 322(5909):1849-51.

Sekido R, Lovell-Badge R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*. 453(7197):930-4.

Sekido R. (2010). SRY: A transcriptional activator of mammalian testis determination. *Int J Biochem Cell Biol*. 42(3):417-20.

Sekiguchi T, Iida H, Fukumura J, Nishimoto T. (2004). Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. *Exp Cell Res*. 300(1):213-22.

Sekiguchi T, Kurihara Y, Fukumura J. (2007). Phosphorylation of threonine 204 of DEAD-box RNA helicase DDX3 by cyclin B/cdc2 in vitro. *Biochem Biophys Res Commun*. 356(3):668-73.

Sengupta T, Cohet N, Morlé F, Bieker JJ. (2009). Distinct modes of gene regulation by a cell-specific transcriptional activator. *Proc Natl Acad Sci USA*. 106(11):4213-8.

Session DR, Lee GS, Wolgemuth DJ. (2001). Characterization of D1Pas1, a mouse autosomal homologue of the human AZFa region DBY, as a nuclear protein in spermatogenic cells. *Fertil Steril*. 76(4):804-11.

Sha F, Mu R, Henderson D, Chen FM. (1999). Self-aggregation of DNA oligomers with XGG trinucleotide repeats: kinetic and atomic force microscopy measurements. *Biophys J*. 77(1):410-23.

Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R. (2005). Core promoter binding by histone-like TAF complexes. *Mol Cell Biol*. 25(1):206-19.

Sheets MD, Ogg SC, Wickens MP. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res*. 18(19):5799-805.

Shevelyov YY, Lavrov SA, Mikhaylova LM, Nurminsky ID, Kulathinal RJ, Egorova KS, Rozovsky YM, Nurminsky DI. (2009). The B-type lamin is required for somatic repression of testis-specific gene clusters. *Proc Natl Acad Sci USA*. 106(9):3282-7.

Shiao MS, Khil P, Camerini-Otero RD, Shiroishi T, Moriwaki K, Yu HT, Long M. (2007). Origins of new male germ-line functions from X-derived autosomal retrogenes in the mouse. *Mol Biol Evol*. 24(10):2242-53.

Shih JW, Tsai TY, Chao CH, Wu Lee YH. (2008). Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein. *Oncogene*. 27(5):700-14.

Shima S, Aigaki T, Nojima T, Yamamoto D. (2007). Identification of trf2 mutants of Drosophila with defects in anterior spiracle eversion. *Arch Insect Biochem Physiol*. 64(4):157-63.

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA*. 100(26):15776-81.

Sikorski TW, Buratowski S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol*. 21(3):344-51.

Simonsson T. (2001). G-quadruplex DNA structures - variations on a theme. *Biol Chem*. 382(4):621-8.

Sinclair A. (2001). Eleven years of sexual discovery. *Genome Biol*. 2(7):REPORTS4017.

Singh GB, Kramer JA, Krawetz SA. (1997). Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res*. 25(7):1419-25.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 423(6942):825-37.

Skandalis A, Frampton M, Seger J, Richards MH. (2010). The adaptive significance of unproductive alternative splicing in primates. *RNA*. 16(10):2014-22.

Smale ST, Baltimore D. (1989). The "initiator" as a transcription control element. *Cell*. 57(1):103-13.

Smale ST. (2001). Core promoters: active contributors to combinatorial gene regulation. *Genes Dev*. 15(19):2503-8.

Smale ST, Kadonaga JT. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*. 72: 449-79.

Smith AD, Sumazin P, Xuan Z, Zhang MQ. (2006). DNA motifs in human and mouse proximal promoters predict tissue-specific expression. *Proc Natl Acad Sci USA*. 103(16):6275-80.

Smith L. (2008). Post-transcriptional regulation of gene expression by alternative 5'-untranslated regions in carcinogenesis. *Biochem Soc Trans*. 36(Pt 4):708-11.

Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, Davenport MP, Mak J. (2010). Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene*. 469(1-2):45-51.

Sobek-Klocke I, Disqué-Kochem C, Ronsiek M, Klocke R, Jockusch H, Breuning A, Ponstingl H, Rojas K, Overhauser J, Eichenlaub-Ritter U. (1997). The human gene ZFP161 on 18p11.21-pter encodes a putative c-myc repressor and is homologous to murine Zfp161 (Chr 17) and Zfp161-rs1 (X Chr). *Genomics*. 43(2):156-64.

Song H, Smolen P, Av-Ron E, Baxter DA, Byrne JH. (2007). Dynamics of a minimal model of interlocked positive and negative feedback loops of transcriptional regulation by cAMP-response element binding proteins. *Biophys J*. 92(10):3407-24.

Sorek R, Shamir R, Ast G. (2004). How prevalent is functional alternative splicing in the human genome? *Trends Genet*. 20(2):68-71.

Soulat D, Bürckstümmer T, Westermayer S, Goncalves A, Bauch A, Stefanovic A, Hantschel O, Bennett KL, Decker T, Superti-Furga G. (2008). The DEAD-box helicase DDX3X is a critical component of the TANK-binding kinase 1-dependent innate immune response. *EMBO J*. 27(15):2135-46.

Speed RM, Vogt P, Köhler MR, Hargreave TB, Chandley AC. (1993). Chromatin condensation behaviour of the Y chromosome in the human testis. I. Evidence for decondensation of distal Yq in germ cells prior to puberty with a switch to Sertoli cells in adults. *Chromosoma*. 102(6):421-7.

Spies N, Nielsen CB, Padgett RA, Burge CB. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell*. 36(2):245-54.

Steger K. (2001). Haploid spermatids exhibit translationally repressed mRNAs. *Anat Embryol (Berl)*. 203(5):323-34.

Stormo GD. (2000). DNA binding sites: representation and discovery. *Bioinformatics*. 16(1):16-23.

Stormo GD, Zhao Y. (2010). Determining the specificity of protein-DNA interactions. *Nat Rev Genet*. 11(11):751-60.

Stothard P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*. 28(6):1102, 1104.

Stouffs K, Tournaye H, Liebaers I, Lissens W. (2009). Male infertility and the involvement of the X chromosome. *Hum Reprod Update*. 15(6):623-37.

Sturm RA, Das G, Herr W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev*. 2(12A):1582-99.

Stutz F, Izaurralde E. (2003). The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol*. 13(6):319-27.

Suganuma T, Workman JL. (2008). Crosstalk among Histone Modifications. *Cell*. 135(4):604-7.

Sugiura K, Muro Y, Nagai Y, Kamimoto T, Wakabayashi T, Ohashi M, Hagiwara M. (1997). Expression cloning and intracellular localization of a human ZF5 homologue. *Biochim Biophys Acta*. 1352(1):23-6.

Sun M, Song L, Li Y, Zhou T, Jope RS. (2008). Identification of an antiapoptotic protein complex at death receptors. *Cell Death Differ*. 15(12):1887-900.

Sun M, Song L, Zhou T, Gillespie GY, Jope RS. (2011). The role of DDX3 in regulating Snail. *Biochim Biophys Acta*. 1813(3):438-47.

Sun T, Xin Z, Jin Z, Wu Y, Gong Y. (2008). Effect of TGF-beta/Smad signaling on sertoli cell and possible mechanism related to complete sertoli cell-only syndrome. *Mol Cell Biochem*. 319(1-2):1-7.

Suomalainen L, Dunkel L, Ketola I, Eriksson M, Erkkilä K, Oksjoki R, Taari K, Heikinheimo M, Pentikäinen V. (2004). Activator protein-1 in human male germ cell apoptosis. *Mol Hum Reprod*. 10(10):743-53.

Suske G. (1999). The Sp-family of transcription factors. *Gene*. 238(2):291-300.

Suzuki MM, Bird A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet.* 9(6):465-76.

Suzuki Y, Sugano S. (2003). Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol Biol*. 221:73-91.

Sverdlov ED, Vinogradova TV. (2010). Core Promoters As an Example of the Effect of Whole-Genome Information on the Evolution of Views on Molecular Mechanisms of Vital Activity. *Molecular Biology*. 44(5):682–692.

Swamynathan SK, Piatigorsky J. (2002). Orientation-dependent influence of an intergenic enhancer on the promoter activity of the divergently transcribed mouse Shsp/alpha B-crystallin and Mkbp/HspB2 genes. *J Biol Chem*. 277(51):49700-6.

Takada S, Lis JT, Zhou S, Tjian R. (2000). A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. *Cell*. 101(5):459-69.

Takai D, Jones PA. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA*. 99(6):3740-5.

Taliaferro JM, Alvarez N, Green RE, Blanchette M, Rio DC. (2011). Evolution of a tissue-specific splicing network. *Genes Dev*. 25(6):608-20.

Tarn WY, Chang TH. (2009). The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol*. 6(1):17-20.

Tassone F, Beilina A, Carosi C, Albertosi S, Bagni C, Li L, Glover K, Bentley D, Hagerman PJ. (2007). Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*. 13(4):555-62.

Tattersall I. (2009). Out of Africa: modern human origins special feature: human origins: out of Africa. *Proc Natl Acad Sci USA*. 106(38):16018-21.

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol*. 14(11):1025-40.

Tazi J, Bird A. (1990). Alternative chromatin structure at CpG islands. *Cell*. 60(6):909-20.

Teif VB. (2010). Predicting gene-regulation functions: lessons from temperate bacteriophages. *Biophys J*. 98(7):1247-56.

Tessadori F, Zeng K, Manders E, Riool M, Jackson D, van Driel R. (2010). Stable S/MAR-based episomal vectors are regulated at the chromatin level. *Chromosome Res*. 18(7):757-75.

Thanaraj TA, Clark F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res*. 29(12):2581-93.

Theisen JW, Lim CY, Kadonaga JT. (2010). Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol Cell Biol*. 30(14):3471-9.

Thomas K, Wu J, Sung DY, Thompson W, Powell M, McCarrey J, Gibbs R, Walker W. (2007). SP1 transcription factors in male germ cell development and differentiation. *Mol Cell Endocrinol*. 270(1-2):1-7.

Thomas MC, Chiang CM. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol*. 41(3):105-78.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 464(7291):1082-6.

Tian B, Hu J, Zhang H, Lutz CS. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*. 33(1):201-12.

Timón M, Beverley PC. (2001). Structural and functional analysis of the human CD45 gene (PTPRC) upstream region: evidence for a functional promoter within the first intron of the gene. *Immunology*. 102(2):180-9.

Tojima Y, Fujimoto A, Delhase M, Chen Y, Hatakeyama S, Nakayama K, Kaneko Y, Nimura Y, Motoyama N, Ikeda K, Karin M, Nakanishi M. (2000). NAK is an IkappaB kinase-activating kinase. *Nature*. 404(6779):778-82.

Tokuhiro K, Miyagawa Y, Yamada S, Hirose M, Ohta H, Nishimune Y, Tanaka H. (2007). The 193-base pair Gsg2 (haspin) promoter region regulates germ cell-specific expression bidirectionally and synchronously. *Biol Reprod*. 76(3):407-14.

Tokusumi Y, Ma Y, Song X, Jacobson RH, Takada S. (2007). The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Mol Cell Biol*. 27(5):1844-58.

Tora L, Timmers HT. (2010). The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends Biochem Sci*. 35(6):309-14.

Torgerson DG, Whitty BR, Singh RS. (2005). Sex-specific functional specialization and the evolutionary rates of essential fertility genes. *J Mol Evol*. 61(5):650-8.

Tóth-Petróczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M. (2008). Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol*. 4(12):e1000243.

Tsukiyama T, Becker PB, Wu C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*. 367(6463):525-532.

Turner JM. (2007). Meiotic sex chromosome inactivation. *Development*. 134(10):1823-31.

Tyler-Smith C. (2008). An evolutionary perspective on Y-chromosomal variation and male infertility. *Int J Androl*. 31(4):376-82.

Tyler-Smith C, Krausz C. (2009). The will-o'-the-wisp of genetics--hunting for the azoospermia factor gene. *N Engl J Med*. 360(9):925-7.

Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics*. 4(12):1920-32.

Umehara T, Kida S, Yamamoto T, Horikoshi M. (1995). Isolation and characterization of a cDNA encoding a new type of human transcription elongation factor S-II. *Gene*. 167(1-2):297-302.

Upadhyaya AB, Khan M, Mou TC, Junker M, Gray DM, DeJong J. (2002). The germ cell-specific transcription factor ALF. Structural properties and stabilization of the TATA-binding protein (TBP)-DNA complex. *J Biol Chem*. 277(37):34208-16.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 10(4):252-63.

Veitia RA. (2010). FOXL2 versus SOX9: a lifelong "battle of the sexes". *Bioessays*. 32(5):375-80.

Ventelä S, Toppari J, Parvinen M. (2003). Intercellular organelle traffic through cytoplasmic bridges in early spermatids of the rat: mechanisms of haploid gene product sharing. *Mol Biol Cell*. 14(7):2768-80.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. (2001). The sequence of the human genome. *Science*. 291(5507):1304-51.

Vergnaud G, Denoeud F. (2000). Minisatellites: mutability and genome architecture. *Genome Res*. 10(7):899-907.

Vernet N, Dennefeld C, Guillou F, Chambon P, Ghyselinck NB, Mark M. (2006). Prepubertal testis development relies on retinoic acid but not rexinoid receptors in Sertoli cells. *EMBO J*. 25(24):5816-25.

Veyrunes F, Waters PD, Miethke P, Rens W, McMillan D, Alsop AE, Grützner F, Deakin JE, Whittington CM, Schatzkamer K, et al. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res*. 18(6):965-73.

Vidal VP, Chaboissier MC, de Rooij DG, Schedl A. (2001). Sox9 induces testis development in XX transgenic mice. *Nat Genet*. 28(3):216-7.

Vinckenbosch N, Dupanloup I, Kaessmann H. (2006). Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci USA*. 103(9):3220-5.

Vincze T, Posfai J, Roberts RJ. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res*. 31(13):3688-91.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res*. 35(Database issue):D88-92.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, Schill WB, Farah S, Ramos C, et al. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet*. 5(7):933-43.

Vogt PH. (2005). Azoospermia factor (AZF) in Yq11: towards a molecular understanding of its function for human male fertility and spermatogenesis. *Reprod Biomed Online*. 10(1):81-93.

Vogt PH. (2006). Genetics: a basic science perspective. In De Jonge CJ, Barratt C. (ed.), *The Sperm Cell: Production, Maturation, Fertilization, Regeneration*. Cambridge University Press, Cambridge, UK, 1. Auflage, pp. 217 - 260.

Vogt PH, Ditton HJ, Kamp C, Zimmer J. (2007). Structure and Function of AZFa Locus in Human Spermatogenesis. In Lau YFC, Chan WY (ed.), *The Y Chromosome and Male Germ Cell Biology in Health and Diseases*. World Scientific Publishing Company, Singapur, Singapur, 1. Auflage, pp 91-126.

Vogt PH, Falcao CL, Hanstein R, Zimmer J. (2008). The AZF proteins. *Int J Androl*. 31(4):383-94.

Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC. (1992). The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science*. 258(5079):52-9.

von der Haar T, Gross JD, Wagner G, McCarthy JE. (2004). The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nat Struct Mol Biol*. 11(6):503-11.

von Haeseler A, Liebers D. (2003). *Molekulare Evolution*. Fischer Taschenbuch Verlag, Frankfurt am Main.

von Moeller F, Barendziak T, Apte K, Goldberg MW, Stick R. (2010). Molecular characterization of Xenopus lamin LIV reveals differences in the lamin composition of sperms in amphibians and mammals. *Nucleus*. 1(1):85-95.

Vong QP, Li Y, Lau YF, Dym M, Rennert OM, Chan WY. (2006). Structural characterization and expression studies of Dby and its homologs in the mouse. *J Androl*. 27(5):653-61.

Wagner S, Hess MA, Ormonde-Hanson P, Malandro J, Hu H, Chen M, Kehrer R, Frodsham M, Schumacher C, Beluch M, et al. (2000): A broad role for the zinc finger protein ZNF202 in human lipid metabolism. *J Biol Chem*. 275(21):15685-90.

Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. (2006). Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res*. 34(Web Server issue):W541-5.

Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K. (2008). DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res*. 36(Database issue):D97-101.

Walker JD, Geissman JW. (2009). 2009 GSA Geologic Time Scale. *GSA Today*. 19(4):60-61.

Walker WH, Habener JF. (1996). Role of transcription factors CREB and CREM in cAMP-regulated transcription during spermatogenesis. *Trends Endocrinol Metab*. 7(4):133-8.

Wallace JA, Felsenfeld G. (2007). We gather together: insulators and genome organization. *Curr Opin Genet Dev*. 17(5):400-7.

Wallis MC, Waters PD, Graves JA. (2008). Sex determination in mammals--before and after the evolution of SRY. *Cell Mol Life Sci*. 65(20):3182-95.

Wang H, Song W, Hu T, Zhang N, Miao S, Zong S, Wang L. (2010 Oct 27. Epub ahead of print). Fank1 interacts with Jab1 and regulates cell apoptosis via the AP-1 pathway. *Cell Mol Life Sci*.

Wang J, Kudoh J, Takayanagi A, Shimizu N. (2005). Novel human BTB/POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161. *Biochem Biophys Res Commun*. 327(2):615-27.

Wang PJ, Page DC. (2002). Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum Mol Genet*. 11(19):2341-6.

Wang PJ. (2004). X chromosomes, retrogenes and their role in male reproduction. *Trends Endocrinol Metab*. 15(2):79-83.

Wang TY, Han ZM, Chai YR, Zhang JH. (2010). A mini review of MAR-binding proteins. *Mol Biol Rep*. 37(7):3553-60.

Waters PD, Wallis MC, Marshall Graves JA. (2007). Mammalian sex - Origin and evolution of the Y chromosome and SRY. *Semin Cell Dev Biol*. 18(3):389-400.

(Mouse Genome Sequencing Consortium), Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*. 420(6915):520-62.

Wegner M. (2010). All purpose Sox: The many roles of Sox proteins in gene expression. *Int J Biochem Cell Biol*. 42(3):381-90.

Wei W, Brennan MD. (2000). Polarity of transcriptional enhancement revealed by an insulator element. *Proc Natl Acad Sci USA*. 97(26):14518-23.

Weirauch MT, Hughes TR. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet*. 26(2):66-74.

Wells DG. (2006). RNA-binding proteins: a lesson in repression. *J Neurosci*. 26(27):7135-8.

West AG, Gaszner M, Felsenfeld G. (2002). Insulators: many functions, many mechanisms. *Genes Dev*. 16(3):271-88.

Whyte DA, Li C, Thomson RB, Nix SL, Zanjani R, Karp SL, Aronson PS, Igarashi P. (1999). Ksp-cadherin gene promoter. I. Characterization and renal epithelial cell-specific activity. *Am J Physiol*. 277(4 Pt 2):F587-98.

Widłak W, Vydra N, Dudaladava V, Scieglińska D, Winiarski B, Krawczyk Z. (2007). The GC-box is critical for high level expression of the testis-specific Hsp70.2/Hst70 gene. *Acta Biochim Pol*. 54(1):107-12.

Wieland M, Hartig JS. (2007). RNA quadruplex-based modulation of gene expression. *Chem Biol*. 14(7):757-63.

Wilkie GS, Dickson KS, Gray NK. (2003). Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends Biochem Sci*. 28(4):182-8.

Wilkinson FH, Park K, Atchison ML. (2006). Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc Natl Acad Sci USA*. 103(51):19296-301.

Wilkinson RD, Steiper ME, Soligo C, Martin RD, Yang Z, Tavaré S. (2011). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Syst Biol*. 60(1):16-31.

Williams A, Flavell RA. (2008). The role of CTCF in regulating nuclear organization. *J Exp Med*. 205(4):747-50.

Williams RR, Hassan-Walker AF, Lavender FL, Morgan M, Faik P, Ragoussis J. (2001). The minisatellite of the GPI/AMF/NLK/MF gene: interspecies conservation and transcriptional activity. *Gene*. 269(1-2):81-92.

Williams TM, Carroll SB. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat Rev Genet*. 10(11):797-804.

Wilson MA, Makova KD. (2009). Evolution and survival on eutherian sex chromosomes. *PLoS Genet*. 5(7):e1000568.

Wimmer R, Kirsch S, Rappold GA, Schempp W. (2005). The evolution of the azoospermia factor region AZFa in higher primates. *Cytogenet Genome Res*. 108(1-3):211-6.

Wingender E. (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform*. 9(4):326-32.

Wong HM, Huppert JL. (2009). Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol Biosyst*. 5(12):1713-9.

Wong SP, Argyros O, Coutelle C, Harbottle RP. (2011). Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage. *Gene Ther*. 18(1):82-7.

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, Bourc'his D, Oakey RJ. (2008). Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev*. 22(9):1141-6.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 20(9):1377-419.

Wray GA. (2007). The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 8(3):206-16.

Wu CI, Davis AW. (1993). Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *Am Nat*. 142(2):187-212.

Wu DY, Yao Z. (2005). Isolation and characterization of the murine Nanog gene promoter. *Cell Res*. 15(5):317-24.

Wu S, Hu YC, Liu H, Shi Y. (2009). Loss of YY1 impacts the heterochromatic state and meiotic double-strand breaks during mouse spermatogenesis. *Mol Cell Biol*. 29(23):6245-56.

Wu Y, Brosh RM Jr. (2010). G-quadruplex nucleic acids and human disease. *FEBS J*. 277(17):3470-88.

Wunderle VM, Critcher R, Ashworth A, Goodfellow PN. (1996). Cloning and characterization of SOX5, a new member of the human SOX gene family. *Genomics*. 36(2):354-8.

Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z. (2007). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res*. 17(6):798-806.

Xia X, Holcik M. (2009). Strong eukaryotic IRESs have weak secondary structure. *PLoS One*. 4(1):e4136.

Xu J, Beyer AR, Walker WH, McGee EA. (2003). Developmental and stage-specific expression of Smad2 and Smad3 in rat testis. *J Androl*. 24(2):192-200.

Yagil G. (2006). DNA tracts composed of only two bases concentrate in gene promoters. *Genomics*. 87(5):591-7.

Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*. 389(1):52-65.

Yang LV, Heng HH, Wan J, Southwood CM, Gow A, Li L. (2003). Alternative promoters and polyadenylation regulate tissue-specific expression of Hemogen isoforms during hematopoiesis and spermatogenesis. *Dev Dyn*. 228(4):606-16.

Yang Q, Jankowsky E. (2005). ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry*. 44(41):13591-601.

Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature*. 419(6907):641-5.

Yedavalli VS, Neuveut C, Chi YH, Kleiman L, Jeang KT. (2004). Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell*. 119(3):381-92.

Yoder AD. (2007). Lemurs. *Curr Biol*. 17(20):R866-8.

Yonaha M, Proudfoot NJ. (1999). Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell*. 3(5):593-600.

Young PP, Mendelson CR. (1997). A GT box element is essential for basal and cyclic adenosine 3',5'-monophosphate regulation of the human surfactant protein A2 gene in alveolar type II cells: evidence for the binding of lung nuclear factors distinct from Sp1. *Mol Endocrinol*. 11(8):1082-93.

Zalmout IS, Sanders WJ, Maclatchy LM, Gunnell GF, Al-Mufarreh YA, Ali MA, Nasser AA, Al-Masari AM, Al-Sobhi SA, Nadhra AO, et al. (2010). New Oligocene primate from Saudi Arabia and the divergence of apes and Old World monkeys. *Nature*. 466(7304):360-4.

Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. (2003). Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res*. 31(5):1375-86.

Zhang H, Lee JY, Tian B. (2005). Biased alternative polyadenylation in human tissues. *Genome Biol*. 6(12):R100.

Zhao C, Meng A. (2005). Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Dev Growth Differ*. 47(4):201-11.

Zhao J, Hyman L, Moore C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*. 63(2):405-45.

Zhao Z, Han L. (2009). CpG islands: algorithms and applications in methylation studies. *Biochem Biophys Res Commun*. 382(4):643-5.

Zhuang Y, Butler B, Hawkins E, Paguio A, Orr L, Wood MG, Wood KV. (2001). A New Age of Enlightenment: New Synthetic *Renilla* Gene and Assay System Increase Expression, Reliability and Sensitivity. *Promega Notes*. 79:6-11.

Zuckerhandl E, Pauling L. (1965). Evolutionary divergence and convergence in proteins. In Bryson V, Vogel HJ (ed.), *Evolving genes and proteins*. Academic Press, Waltham, USA, pp. 97-166.

Zuker M, Mathews DH, Turner DH. (1999). Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. In Barciszewski J, Clark BFC (ed.), *RNA Biochemistry and Biotechnology*. NATO ASI Series, Kluwer Academic Publishers, Dordrecht, NL, pp. 11-43.

Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31(13):3406-15.

7. Abbildungsverzeichnis

Abb. 1-1	Schematische Darstellung der konservierten Motive in DEAD-Box RNA-Helikasen	1
Abb. 1-2	Lokalisation des DDX3X Proteins in einem Human Hodengewebsschnitt	3
Abb. 1-3	Lokalisation des DDX3Y Proteins in einem Human Hodengewebsschnitt	6
Abb. 1-4	Schematische Darstellung des Human Y-Chromosoms und der AZFa Deletion	7
Abb. 1-5	Schematischer Wirbeltierstammbaum mit evolutionärer Zeittafel	10
Abb. 1-6	Schematischer Stammbaum der Primaten	13
Abb. 1-7	Human Geschlechtschromosomen im schematischen Vergleich	14
Abb. 1-8	Y-Chromosomen fünf verschiedener Hominiden in Bezug zur AZFa Lokalisation	16
Abb. 1-9	Schematischer Vergleich des Human und Schimpansen Y-Chromosoms	17
Abb. 1-10	Schematische Darstellung verschiedener Core-Promotermotive und deren Erkennung durch Untereinheiten des Transkriptionskomplexes	20
Abb. 1-11	Eukaryotische Promoterstruktur und Promoterfunktion	27
Abb. 1-12	Chromatindomänen und regulative Elemente im Schema	30
Abb. 1-13	Übertragung entfernter genspezifischer TF-Signale auf die Transkriptionsmaschinerie durch den verformbaren Mediator-Komplex	31
Abb. 1-14	Vereinfachte, schematische Darstellung der Translationsinitiation und der Ringbildung zwischen 5´UTR und 3´UTR	35
Abb. 1-15	Schematische Darstellung möglicher RNA-Sekundärstrukturen	37
Abb. 1-16	Darstellung alternativer Möglichkeiten der Translationsinitiation	38
Abb. 1-17	Schematische Darstellung alternativer 5´- und 3´UTR-Sequenzen	39
Abb. 3-1	*Beta-Actin* Kontrolle am Beispiel *Macaca mulatta*	58
Abb. 3-2	Prinzip der Sequenzierung	65
Abb. 3-3	Mögliches Bandenmuster eines sequenzierten Fragmentes	66
Abb. 3-4	Prinzip der Sequenzierung mit Fluoreszenz-markierten ddNTPs	66
Abb. 3-5	Schematische Darstellung des Luziferase T7-Kontrollplasmid	69
Abb. 3-6	Schematische Darstellung der pUC18 Vektorkarte	70
Abb. 3-7	Schematische Darstellung des ATG-Translationskontext in den *in-vitro* pUC-18 Plasmiden	70
Abb. 3-8	Struktur des modifizierten Substrates 5´Fluoroluziferin aus dem ONE-GLO Luciferase Assay System	72
Abb. 3-9	Von Firefly- und Renilla-Luziferase katalysierte Biolumineszenzreaktionen	75

Abb. 4-1	RT-PCR Analyse zur Kartierung der absoluten Grenze der *DDX3Y* TSS-I Region	81
Abb. 4-2	Genomische Sequenz des *DDX3Y* Genlocus von Position 53.700 bis 55.190 (Ende Exon1) im BAC 475I1	82
Abb. 4-3	Core-Promoter Motive in der *MSY2*-Repeatsequenz stromaufwärts des *DDX3Y* Gens	85
Abb. 4-4	RT-PCR Experimente zur Eingrenzung der *DDX3Y* Transkriptionsstartstelle im *MSY2* Repeat	87
Abb. 4-5	*MSY2* Repeat-Sequenz (Position 53.854 – 54.262 in BAC 475I1) mit eingezeichneten Primersequenzen zur RT-PCR Analyse	88
Abb. 4-6	RT-PCR Analyse zur Bestimmung der *DDX3Y* Transkriptionsinitiation in *MSY2*	89
Abb. 4-7	RT-PCR Analysen zur Transkriptinitiation in der *DDX3Y* Exon-T Region	91
Abb. 4-8	Graphische Zusammenfassung der *DDX3Y* Transkriptionsinitiationsstellen	91
Abb. 4-9	Genomische Sequenz des *DDX3X* Genlocus von Position 29.785 bis 30.774 (Ende Exon1) im BAC RP13-13A3	96
Abb. 4-10	RT-PCR Analyse der *DDX3X* TSS-I Region	97
Abb. 4-11	RT-PCR Analyse zur Eingrenzung der selektiv genutzten *DDX3X* T/B-TSS-II	98
Abb. 4-12	RT-PCR Analyse *DDX3X* T-TSS-III Transkripte	100
Abb. 4-13	Graphische Darstellung der drei *DDX3X* Transkriptionsinitiationsstellen	101
Abb. 4-14	Vergleich der genomischen *DDX3Y* und *DDX3X* 5´ Sequenzen	104
Abb. 4-15	Genomisches *DDX3Y* Sequenzalignment zwischen Mensch (Hsap), Schimpanse (Ptro), Rhesusaffe (Mmul) und Weißbüschelaffe (Cjac)	107
Abb. 4-16	RT-PCR Analyse von *Pan troglodytes DDX3Y* TSS-I Region	109
Abb. 4-17	RT-PCR Analyse *Macaca mulatta DDX3Y* TSS-I Region	110
Abb. 4-18	RT-PCR Analyse *Callithrix jacchus DDX3Y* TSS-I Region	110
Abb. 4-19	*Callithrix jacchus DDX3Y* RT-PCR Analyse im Bereich von TSS-I-(T)	111
Abb. 4-20	Sequenzalignment der *MSY2* Repeateinheit 1 zwischen *Homo*, *Pan*, *Macaca* und *Callithrix*	113
Abb. 4-21	Sequenzalignment der *MSY2* Repeateinheit 2 zwischen *Homo*, *Pan*, *Macaca* und *Callithrix*	113
Abb. 4-22	RT-PCR Analyse *Pan troglodytes DDX3Y* Transkriptstart im Bereich *MSY2* und Exon-T	114
Abb. 4-23	RT-PCR Analyse *Macaca mulatta DDX3Y* Transkriptstart im Bereich *MSY2*	116
Abb. 4-24	*Macaca mulatta* RT-PCR im Bereich von *DDX3Y* Exon-T	117
Abb. 4-25	*Callithrix jacchus DDX3Y* RT-PCR im Bereich von *MSY2* und Exon-T	117
Abb. 4-26	Vergleich der *DDX3Y* Expression im Bereich von *MSY2* und Exon-T im Hodengewebe von Mensch, Schimpanse, Rhesusaffe und Weißbüschelaffe	119
Abb. 4-27	Genomisches *DDX3Y/ Ddx3y* Sequenzalignment zw. Mensch (Hsap), Rind (Btau) und Maus (Mmus)	121

Abb. 4-28	*Ddx3y* RT-PCR Analyse *Mus musculus*	123
Abb. 4-29	RT-PCR Analyse Rind *Ddx3y*	125
Abb. 4-30	RT-PCR Analyse Rind *Ddx3y*, Exon-T und *MSY2* Region	127
Abb. 4-31	Graphische Zusammenfassung der Rind *Ddx3y* Transkriptionsinitiationsstellen	129
Abb. 4-32	Genomisches *DDX3X* Sequenzalignment zwischen Mensch (Hsap), Schimpanse (Ptro), Rhesusaffe (Mmul) und Weißbüschelaffe (Cjac)	131
Abb. 4-33	Schimpansen (Ptro) *DDX3X* TSS-I RT-PCR Analyse	133
Abb. 4-34	*Macaca* (Mmul) *DDX3X* TSS-I RT-PCR Analyse	134
Abb. 4-35	*Callithrix* (Cjac) *DDX3X* TSS-I RT-PCR Analyse	134
Abb. 4-36	Schimpanse (Ptro) *DDX3X* T-TSS-III RT-PCR Analyse	135
Abb. 4-37	*Macaca* (Mmul) *DDX3X* T-TSS-III RT-PCR Analyse	136
Abb. 4-38	*Callithrix* (Cjac) *DDX3X* T-TSS-III RT-PCR Analyse	137
Abb. 4-39	Schematische Darstellung des *in-silico* vorhergesagten *Macaca mulatta DDX3X* Transkriptes „XM_001095294“	138
Abb. 4-40	RT-PCR Analyse des vorhergesagten *Macaca mulatta DDX3X* Transkriptes „XM_001095294“	139
Abb. 4-41	Genomisches *DDX3X/ Ddx3x* Sequenzalignment zwischen Mensch (Hsap), Rind (Btau) und Maus (Mmus)	140
Abb. 4-42	keine Abbildung	
Abb. 4-43	RT-PCR Analyse Maus *Ddx3x*	143
Abb. 4-44	RT-PCR Analyse Rind *Ddx3x*	145
Abb. 4-45	Graphischer VISTA-Plot des Alignments zwischen *Hsap DDX3Y* und *DDX3X*	151
Abb. 4-46	Graphischer VISTA-Plot des Alignments zwischen Hsap *DDX3X* und Mdom *Ddx3*	158
Abb. 4-47	Graphische Auswertung der CpG-Islands und DNA-Signaturen in den *DDX3Y* und *DDX3X* Promoterregionen	159
Abb. 4-48	Graphische Zusammenfassung der *DDX3Y* und *DDX3X* TF-Analyse	163
Abb. 4-49	Graphische Darstellung der Abstände zwischen konservierten TFBSs und gemeinsamer TF-Module in der Human *DDX3Y* und *DDX3X* Promoterregion	180
Abb. 4-50	mVISTA-, cTFBS- und CpG-Island-Vergleich der Human, Schimpanse, *Macaca* und *Callithrix DDX3Y* Promoterregionen	184
Abb. 4-51	mVISTA-, cTFBS- und CpG-Island-Vergleich der Human *DDX3Y* und Rind und Maus *Ddx3y* Promoterregionen	191
Abb. 4-52	mVISTA-Vergleich der Human, Schimpanse, *Macaca*, *Callithrix*, Rind und Maus *DDX3X/ Ddx3x* Promoterregionen	199
Abb. 4-53	Schematische Darstellung der Human Gonosomen im Bereich der *DDX3Y* und *DDX3X* Genloci	203

Abb. 4-54 RT-PCR Analyse des *DDX3X* Exon17 Splicemusters im Hodengewebe von Schimpanse und Human 205

Abb. 4-55 Karte der MAR-Vorhersage mit MARFinder für den *DDX3Y* Locus 207

Abb. 4-56 Karte der MAR-Vorhersage mit MARFinder für den *DDX3X* Locus 208

Abb. 4-57 RT-PCR Analyse *DDX3X* Intron1 Gen 212

Abb. 4-58 Graphische Darstellung der *DDX3X* Promoterkonstrukte 215

Abb. 4-59 Auswertung des Human *DDX3X* Promoterregion DLR-Assays 216

Abb. 4-60 Graphische Darstellung der partiellen *DDX3X* Promoterkonstrukte 220

Abb. 4-61 Auswertung der partiellen *DDX3X* Promoterkonstrukte I 221

Abb. 4-62 Auswertung der partiellen *DDX3X* Promoterkonstrukte II 224

Abb. 4-63 Auswertung der partiellen *DDX3X* Promoterkonstrukte III 226

Abb. 4-64 Vergleich der Human *DDX3Y MSY2* und *DDX3X MSY2-X* Sequenzen 230

Abb. 4-65 Schematische Darstellung der *DDX3Y MSY2* Konstrukte 231

Abb. 4-66 Auswertung des *DDX3Y MSY2* DLR-Assays 234

Abb. 4-67 Darstellung der eingeführten Mutationen in die *MSY2* und *MSY2-X* Sequenzen 236

Abb. 4-68 Auswertung der eingeführten Mutationen in die *DDX3Y MSY2* Repeats 237

Abb. 4-69 Schematische Darstellung der *DDX3X MSY2-X* Konstrukte 239

Abb. 4-70 Auswertung des *DDX3X MSY2-X* DLR-Assays 240

Abb. 4-71 Auswertung der *DDX3X MSY2-X* Mutationskonstrukte 242

Abb. 4-72 Vergleich der genomischen *DDX3Y* und *DDX3X* Sequenzen in den 5´UTR-Regionen 245

Abb. 4-73 Entstehung von *DDX3X* uORF2 durch potentielle Duplikation in der Human *DDX3X* 5´UTR Sequenz 246

Abb. 4-74 Schematische Darstellung potentiell *cis*-regulativer Elemente in der *DDX3X* 5´UTR 253

Abb. 4-75 Sekundärstrukturvorhersage für *DDX3Y* 5´UTR (Ta-1b Variante) mit Start in BAC-Position 54.051 im *MSY2-2* Repeat (T-TSS-II) 257

Abb. 4-76 Sekundärstrukturvorhersage für *DDX3X* 5´UTR mit Start in BAC-Position 29.825 (T-TSS-III) 258

Abb. 4-77 CART-Modell zur Einordnung humaner 5´UTR Sequenzen 259

Abb. 4-78 Schematische Darstellung der verschiedenen Human *DDX3X* 5´UTR Konstruktvarianten 267

Abb. 4-79 Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays 270

Abb. 4-80 Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays mit Mutation in uORF1 272

Abb. 4-81 Ergebnisse des *DDX3X* 5´UTR *in-vitro* Translations-Assays in non-Human Primaten 274

Abb. 5-1 Mögliche alternative Chromatinstruktur nach lokaler Verdrängung eines Nukleosoms durch einen GAF-Komplex 296

Abb. 5-2 Struktur und Funktion der Kernlamina 299

Abb. 5-3 Schematische Darstellung potentieller G-Quadruplexsequenzen in DNA- und RNA-Molekülen 305

Abb. 5-4 Potentieller Einfluß von G-Quadruplexsequenzen auf die Translationsfähigkeit einer mRNA 313

Abb. A-1 mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom I 376

Abb. A-2 mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom II 377

Abb. A-3 mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom III 378

Abb. A-4 Genomisches *DDX3X/Ddx3x* Sequenzalignment von *MSY2-X* bis Exon1 379

Abb. A-5 Potentielle uATGs in DDX3X Promoterkonstrukt I (2,44 kb) 382

8. Tabellenverzeichnis

Tab. 1-1	MSY Gene und Genfamilien in Human und Schimpansen Y-Chromosom	18
Tab. 1-2	Liste bekannter Core-Promotermotive	21
Tab. 2-1	Geräte und Zubehör	41
Tab. 2-2	Verbrauchsmaterialien	42
Tab. 2-3	Chemikalien	43
Tab. 2-4	Fertigsysteme	44
Tab. 2-5	Enzyme	44
Tab. 2-6	Medien und Zusätze für Zellkulturen	48
Tab. 2-7	Nukleotide	50
Tab. 2-8	DNA-Größenstandards	50
Tab. 2-9	Verwendete kommerzielle Vektoren und modifizierte Vektoren	50
Tab. 2-10	Verwendete Internet-basierte Datenbanken und Programme	54
Tab. 3-1	Pipettierschema für die Sequenzierungsreaktion der pGL4-Konstrukte	67
Tab. 3-2	PCR-Programm für die Sequenzierungsreaktion mit dem „BigDye Terminator v1.1 Cycle Sequencing Kit“	67
Tab. 4-1	Position potentieller Initiator Core-Promotermotive im Bereich des ubiquitären *DDX3Y* Transkriptstartes in Exon1 (TSS-I Region)	79
Tab. 4-2	Übersicht der Inr-Motive in den vier *MSY2* Repeateinheiten	85
Tab. 4-3	Zusammenfassung der *MSY2* RT-PCR Experimente und Angabe der PCR-Amplifikatgrößen	88
Tab. 4-4	Position potentieller Initiator Core-Promotermotive für den *DDX3X* Transkriptstart	93
Tab. 4-5	Positionen potentieller Initiator-Motive im Bereich des ubiquitären Maus *Ddx3y* Transkriptstartes in Exon1 (TSS-I Region)	122
Tab. 4-6	Positionen potentieller Initiator-Motive im Bereich des ubiquitären Maus *Ddx3x* Transkriptstartes in Exon1 (TSS-I Region)	142
Tab. 4-7	Liste der mit mVISTA ermittelten konservierten Regionen (KRiH) zwischen der *DDX3Y* und *DDX3X* Promoterregion	149
Tab. 4-8	Auflistung aller homopolymeren und binären Sequenzabschnitte in den 2,5 kb Promoterregionen von Human *DDX3Y* und *DDX3X*	153
Tab. 4-9	Vergleich der pot. CpG-Islands in den Promoterregionen von Human *DDX3Y* und *DDX3X*	155
Tab. 4-10	Konservierte TFBS (cTFBS) in Human *DDX3Y* und *DDX3X*	161
Tab. 4-11	Über- und Unterrepräsentierte TFBS in der *DDX3Y* Promoterregion	171

Tab. 4-12 Über- und Unterrepräsentierte TFBS in der *DDX3X* Promoterregion 172

Tab. 4-13 Vergleich der über- und unterrepräsentierten TF-Familien in den *DDX3Y* und *DDX3X* Promoterregionen 173

Tab. 4-14 Auflistung der potentiellen TF-Module in der Human *DDX3Y* Promotersequenz 177

Tab. 4-15 Auflistung der potentiellen TF-Module in der Human *DDX3X* Promotersequenz 178

Tab. 4-16 Liste der mit mVISTA ermittelten konservierten Sequenzabschnitte zwischen der Human *DDX3Y* und den Primaten *DDX3Y* Promoterregionen von Schimpanse, *Macaca* und *Callithrix* 182

Tab. 4-17 Vertikaler Vergleich der pot. CpG-Islands in den *DDX3Y*/*Ddx3y* Promoterregionen von Mensch, Schimpanse, Rhesusaffe, *Callithrix*, Rind und Maus 183

Tab. 4-18 Konservierte TFBS zwischen der Human und non-Human Primaten *DDX3Y* Promoterregion 185

Tab. 4-19 Liste der mit mVISTA ermittelten konservierten Regionen zwischen der Human *DDX3Y* und Rind *Ddx3y* Promoterregion 190

Tab. 4-20 Konservierte TFBS in Human *DDX3Y* und Rind *Ddx3y* 192

Tab. 4-21 Liste der mit mVISTA ermittelten konservierten Regionen zwischen der Human *DDX3Y* und Maus *Ddx3y* Promoterregion 194

Tab. 4-22 Konservierte TFBS in Human *DDX3Y* und Maus *Ddx3y* 195

Tab. 4-23 Liste der mit mVISTA ermittelten konservierten Sequenzabschnitte zwischen den Human, Non-Human Primaten, Rind und Maus *DDX3X/Ddx3x* Promoterregionen 197

Tab. 4-24 Vertikaler Vergleich der pot. CpG-Islands in den *DDX3X*/*Ddx3x* Promoterregionen von Mensch, Schimpanse, Rhesusaffe, *Callithrix*, Rind und Maus 198

Tab. 4-25 Konservierte TFBS zwischen der Human *DDX3X* und den Non-Human Primaten, Rind und Maus *DDX3X/Ddx3x* Promoterregionen 200

Tab. 4-26 Regionen im *DDX3Y* Locus mit potentiellen S/MAR-Elementen 206

Tab. 4-27 Liste der *DDX3X* Promoterkonstrukte 214

Tab. 4-28 Absolute und relative Meßwerte der *DDX3X* Promoterregion DLR-Analyse 217

Tab. 4-29 Liste der partiellen *DDX3X* Promoterkonstrukte 219

Tab. 4-30 Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion I DLR-Analyse 222

Tab. 4-31 Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion II DLR-Analyse 225

Tab. 4-32 Absolute und relative Meßwerte der partiellen *DDX3X* Promoterregion III DLR-Analyse 227

Tab. 4-33 Liste der *DDX3Y MSY2* Konstrukte 232

Tab. 4-34 Absolute und relative Meßwerte der *DDX3Y MSY2* Luziferase-Konstrukte 233

Tab. 4-35 Liste der *DDX3Y MSY2* Konstrukte mit eingeführter Mutation 237

Tab. 4-36 Absolute und relative Meßwerte der *DDX3Y MSY2* Mutationskonstrukte 238

Tab. 4-37 Liste der *DDX3X MSY2-X* Konstrukte 239

Tab. 4-38	Absolute und relative Meßwerte der *DDX3X MSY2-X* Konstrukte	240
Tab. 4-39	Liste der *DDX3X MSY2-X* Konstrukte mit eingeführter Mutation	241
Tab. 4-40	Absolute und relative Meßwerte der *DDX3X MSY2-X* Mutationskonstrukte	242
Tab. 4-41	Vergleich von *DDX3X* uORF1 und *DDX3Y* uORF5 ATG-Kontext zu Kozak-Sequenz, TISU-Element und YY1 TF-Bindesequenz	247
Tab. 4-42	Vergleich binärer „Sequenz-Trakte“ zwischen den *DDX3X* und *DDX3Y* 5´UTR-Varianten	248
Tab. 4-43	Kartierung potentieller G-Quadruplexsequenzen in den Human *DDX3X*/*DDX3Y* 5´UTRs	250
Tab. 4-44	Vergleich der freien Energien ΔG, der 5´UTR-Länge und des (C,G)-Gehaltes der verschiedenen *DDX3Y* und *DDX3X* 5´UTR-Varianten	254
Tab. 4-45	Vertikaler und lateraler (horizontaler) Vergleich des Human *DDX3X* uORF1 und *DDX3Y* uORF5 in genomischen Säugersequenzen	261
Tab. 4-46	Vertikaler Vergleich von Human *DDX3X* uORF2 in verschiedenen Säugerspezies	263
Tab. 4-47	Vergleich der potentiellen Human *DDX3X* 5´UTR G-Quadruplex Sequenzen in *DDX3X* 5´UTRs von non-Human Primaten	264
Tab. 4-48	Zusammenfassung und Speziesvergleich potentieller *cis*-Faktoren in der Human *DDX3X* 5´UTR	265
Tab. 4-49	Beschreibung der verschiedenen *DDX3X* 5´UTR Konstrukte für den *in-vitro* Translations-Assay	268
Tab. 4-50	Absolute und relative Werte der *in-vitro* Messung	271
Tab. 4-51	Beschreibung der verschiedenen non-Human Primaten *DDX3X* 5´UTR Konstrukte für den *in-vitro* Translations-Assay	273
Tab. 4-52	Absolute und relative Werte des *DDX3X* 5´UTR *in-vitro* Translationsansatzes mit non-Human Primaten-Konstrukten	275
Tab. A-1	Liste der Nukleotide in der Schreibweise des „Ambiguity Code“	383
Tab. A-2	Vergleich *DDX3Y* BAC-Positionen mit Y-Chromosom Referenzsequenz	383
Tab. A-3	Vergleich *DDX3X* BAC-Positionen mit X-Chromosom Referenzsequenz	384
Tab. A-4	Human *DDX3Y* Exon-Intronstruktur in BAC-Klon „475I1“ (GenBank Acc.No. AC004474.1)	385
Tab. A-5	Schimpanse *DDX3Y* Exon-Intronstruktur in BAC-Klon „CH251-128L22“ (GenBank Acc.No. AC146254.2) revers-komplementär	386
Tab. A-6	*Macaca DDX3Y* Exon-Intronstruktur in BAC-Klon „CH250-541A11“ (GenBank Acc.No. AC213321.3) revers-komplementär	387
Tab. A-7	*Callithrix DDX3Y* Exon-Intronstruktur in BAC-Klon „CH259-161K20“ (GenBank Acc.No. AC225609.4) revers-komplementär	388
Tab. A-8	Rind *Ddx3y* Exon-Intronstruktur in BAC-Klon „CH240-56A13“ (GenBank Acc.No. AC217167.7) revers-komplementär	389
Tab. A-9	Maus *Ddx3y* Exon-Intronstruktur in BAC-Klon „RP24-208N6“ (GenBank Acc.No. AC145393.4) revers-komplementär	390

Tab. A-10	Human *DDX3X* Exon-Intronstruktur in BAC-Klon „RP13-13A3“ (GenBank Acc.No. AL391647.16)	391
Tab. A-11	Schimpanse *DDX3X* Exon-Intronstruktur in BAC-Klon „CH251-3A20“ (GenBank Acc.No. AC183643.2)	392
Tab. A-12	*Macaca DDX3X* Exon-Intronstruktur in Ensembl *Macaca* X-Chromosom (Ensembl Assembly MMUL 1.0, Datenbankversion 62.10p ≙ GenBank Acc.No. NC_007878.1)	393
Tab. A-13	*Callithrix DDX3X* Exon-Intronstruktur in BAC-Klon „CH259-42H12“ (GenBank Acc.No. AC193908.1) revers-komplementär	394
Tab. A-14	*Rind DDX3X* Exon-Intronstruktur in genomischen Contig „004.81.scaffold1“ (GenBank Acc.No. NW_001501904.1) revers-komplementär	395
Tab. A-15	Maus *DDX3X* Exon-Intronstruktur in BAC-Klon „RP23-158L19“ (GenBank Acc.No. AL833805.11)	396
Tab. A-16	Liste der potentiellen TF-Bindestellen in Human *DDX3Y* Promotersequenz	CD
Tab. A-17	Liste der potentiellen TF-Bindestellen in Human *DDX3X* Promotersequenz	CD
Tab. A-18	Liste der potentiellen TF-Bindestellen in Schimpanse *DDX3Y* Promotersequenz	CD
Tab. A-19	Liste der potentiellen TF-Bindestellen in *Macaca DDX3Y* Promotersequenz	CD
Tab. A-20	Liste der potentiellen TF-Bindestellen in *Callithrix DDX3Y* Promotersequenz	CD
Tab. A-21	Liste der potentiellen TF-Bindestellen in Rind *Ddx3y* Promotersequenz	CD
Tab. A-22	Liste der potentiellen TF-Bindestellen in Maus *Ddx3y* Promotersequenz	CD
Tab. A-23	Liste der potentiellen TF-Bindestellen in Schimpanse *DDX3X* Promotersequenz	CD
Tab. A-24	Liste der potentiellen TF-Bindestellen in *Macaca DDX3X* Promotersequenz	CD
Tab. A-25	Liste der potentiellen TF-Bindestellen in *Callithrix DDX3X* Promotersequenz	CD
Tab. A-26	Liste der potentiellen TF-Bindestellen in Rind *Ddx3x* Promotersequenz	CD
Tab. A-27	Liste der potentiellen TF-Bindestellen in Maus *Ddx3x* Promotersequenz	CD
Tab. A-28	Liste der potentiellen GAGA TF-Bindestellen 3´ zu Human *DDX3Y* Genlocus in pot. S/MAR-Element Nr.8 (Abb. 4-55)	397
Tab. A-29	Rohdaten *DDX3X* Promoter-Luziferaseassay (siehe Tab. 4-28)	CD
Tab. A-30	Rohdaten *DDX3X* Promoter-Luziferaseassay mit partiellen Promoterkonstrukten (siehe Tab. 4-30)	CD
Tab. A-31	Rohdaten *DDX3Y MSY2*-Luziferaseassay (siehe Tab. 4-34)	CD
Tab. A-32	Rohdaten *DDX3Y MSY2*-Mutations-Luziferaseassay (siehe Tab. 4-36)	CD
Tab. A-33	Rohdaten *DDX3X MSY2*-X Luziferaseassay (siehe Tab. 4-38)	CD
Tab. A-34	Rohdaten *DDX3X MSY2*-X Mutations-Luziferaseassay (siehe Tab. 4-40)	CD
Tab. A-35	Rohdaten *DDX3X MSY2*-X Mutations-Luziferaseassay (siehe Tab. 4-40)	CD

Tab. A-36	Rohdaten Human *DDX3X in-vitro* Translationsassay (siehe Tab. 4-50)	CD
Tab. A-37	Rohdaten Primaten *DDX3X in-vitro* Translationsassay (siehe Tab. 4-52)	CD
Tab. A-38	Primersequenzen zur RT-PCR Analyse von *DDX3Y* Transkripten	399
Tab. A-39	Primersequenzen zur RT-PCR Analyse von *DDX3X* Transkripten	401
Tab. A-40	Primersequenzen zur cDNA-Kontrolle (Beta-Actin) per RT-PCR-Analyse	402
Tab. A-41	Primersequenzen zur Insertkontrolle per PCR-Analyse und Sequenzierung	402
Tab. A-42	Primersequenzen zur Klonierung der Human *DDX3X* Promoterregion und der *MSY2*/ *MSY2-X* Sequenzen	403
Tab. A-43	Primersequenzen zur Klonierung der Human und Primaten *DDX3X* 5´UTR-Sequenzen	404

9. Abkürzungsverzeichnis

A	Adenosin
Abb.	Abbildung
Acc.No.	Identifikationsnummer der Gendatenbank
A. dest.	Aqua destillata
AnP	Antarctic Phosphatase
AS	Aminosäure
ATG	Translationsstartcodon
AZF	Azoospermia Faktor
BAC	Bacterial Artificial Chromosome
Bgl	*Bacillus globigii*
BLAST	Basic Local Alignment Search Tool
bp	Basenpaar(e)
BRE	TFIIB recognition element
Btau	*Bos taurus*
BTB/POZ	Broad complex, tamtrack, bric-à-brac/poxvirus and zinc finger Proteindomäne
C	Cytidin
cAMP	zyklisches Adenosinmonophosphat
CART	Classification and regression tree
CB	Chromatoid body
CBP	CREB-binding Protein
cDNA	complementary DNA
CDS	coding sequence
CGI	CpG-Island
Cjac	*Callithrix jacchus*
CNS	conserved non-coding sequence
CpG	CpG-Dinukleotid
CRE	cAMP response element
CREB	CRE-binding Protein
CRM	*cis*-regulatives Modul
cTFBS	konservierte Transkriptionsfaktorbindestelle
Da	Dalton
ddH_2O	doppelt destilliertes Wasser
DDX3X	DEAD-box protein 3, X-chromosomal
DDX3Y	DEAD-box protein 3, Y-chromosomal
DEAD	Asp-Glu-Ala-Asp Motiv
DLR	Dual-Luciferase Reporter Assay
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
dNTP	2´-Desoxynukleosid-5´-triphosphat
ddNTP	2´,3´-Didesoxynukleosid-5´-triphosphat
DPE	Downstream Promoter Element
E. coli	*Escherichia coli*
ed.	Herausgeber (= editor)
eIF	Eukaryotic Initiation Factor
et al.	und andere
EtBr	Ethidiumbromid
EtOH	Ethanol
Exon-T	Exon-Testis
F	forward (vorwärts)
g	Erdbeschleunigung

G	Guanosin
Ggor	*Gorilla gorilla*
GRS	Guanin-rich sequence
GTF	General Transcription Factor
H3	Histon 3
HAT	Histon Acetyltransferase
HERV	Human Endogenous Retrovirus
Hind	*Haemophilus influenzae*
HLH	Helix-Loop-Helix Proteindomäne
HMG	High Mobility Group Proteindomäne
HPA	Human Protein Atlas
HPLP	High Performance Liquid Chromatography
Hsap	*Homo sapiens*
HSV-TK	Herpes simplex Virus Thymidinkinase
indel	insertion and deletion
Inr	Initiator
Intron-T	Intron-Testis
IRES	Internal Ribosome Entry Site
ITAF	IRES-specific *trans*-acting Factor
K_A	nicht-synonyme Substitution
kb	Kilobase
Kpn	*Klebsiella pneumoniae*
KRiH	konservierte Region in Human
K_S	synonyme Substitution
K/T Grenze	Kreide/Tertiär Grenze
Luc	Luziferasegen
LB	Wachstumsmedium nach Luria und Bertani
LCA	last common ancestor
LINE	long interspersed nuclear elements
MARBP	Matrix Attachment Region-binding Protein
Mb	Megabase
MCS	Multiple-Cloning Site
Mdom	*Monodelphis domestica*
Mmul	*Macaca mulatta*
Mmur	*Microcebus murinus*
Mmus	*Mus musculus*
mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein
MSCI	meiotic sex chromosome inactivation
MSY	male-specific region of the Y-chromosome
MSY2	Y-spezifischer Minisatellit Nr. 2
MSY2-X	*MSY2* homologe Sequenz auf X-Chromosom
MTE	Motif Ten Element
MYA	million years ago
NAHR	Non-allelic Homologous Recombination
NCBI	National Center for Biotechnology Information
NDR	Nucleosome-depleted Region
NES	Nucleosome excluding Sequence
NF-κB	Nuclear-Factor kappa B
NLS	Nukleus Lokalisationssignal
NPS	Nucleosome Positioning Sequence
NRY	non-recombining portion of the Y
OD	optische Dichte
OMIM	Online Mendelian Inheritance in Man

ORF	open reading frame (offener Leserahmen)
ORI	Origin of Replication
PABP	Poly(A)-binding Protein
PAR	Pseudoautosomale Region
PARS	Polypurin (A)-reiche Sequenz
PAS	Polyadenylierungsstelle
PBS	phosphate buffered saline
PCAF	p300/CBP-associated factor
PCR	Polymerasekettenreaktion
PIC	Präinitiationskomplex
PMSC	postmeiotic sex chromatin
Ppyg	*Pongo pygmaeus*
PQS	putative quadruplex forming sequence
Ptro	*Pan troglodytes*
Py	Pyrimidin
QGRS	Quadruplex forming G-rich Sequence
R	reverse (rückwärts)
RACE	Rapid Amplification of cDNA ends
RISC	RNA-induced silencing complex
RLU	relativ light unit
RNA	Ribonukleinsäure
RNAPII	RNA-Polymerase II
RT-PCR	Reverse Transcription PCR
SCO	Sertoli Cell-only
SINE	short interspersed nuclear elements
S/MAR	Scaffold/Matrix Attachment Region
SNP	Single Nucleotide Polymorphism
SOX	SRY-related HMG box
SRY	Sex-determining region Y Protein
SSR	Simple Sequence Repeat
SV40	Simian Virus 40
T	Thymidin
Tab.	Tabelle
TAF	TBP-associated Factor
Taq	*Thermophilus aquaticus*
TATA	TATA-Box
TBP	TATA-Box binding Protein
TF	Transkriptionsfaktor
TFII	GTF für RNAPII
TFBS	Transkriptionsfaktorbindestelle
TNR	Trinukleotidrepeat
TOP	Terminal Oligopyrimidin Tract
TRF	TBP-related Factor
TSS	Transkriptionsstartstelle
U	Units
U	Uracil
uATG	upstream Translationsstartcodon
uORF	upstream ORF
UTR	untranslatierte Region
UV	ultraviolettes Licht
WWW	World Wide Web
XAR	X added Region
XCR	X conserved Region
Xp	kurzer Arm des X-Chromosoms, petit (franz. klein)

Xq	langer Arm des X-Chromosoms, q folgt auf p
YAR	Y added Region
ZDR	Z-DNA forming Region
5´	stromaufwärts (upstream)
3´	stromabwärts (downstream)
µl	Mikroliter
µg	Mikrogramm
♂	männlich
♀	weiblich

Für Mengen-, Größen-, Zeit- und andere physikalischen Angaben wurden SI- oder gesetzliche Einheiten benutzt. Aminosäuren wurden mit den international anerkannten Ein- und Dreibuchstaben-Codes dargestellt.

10. Anhang

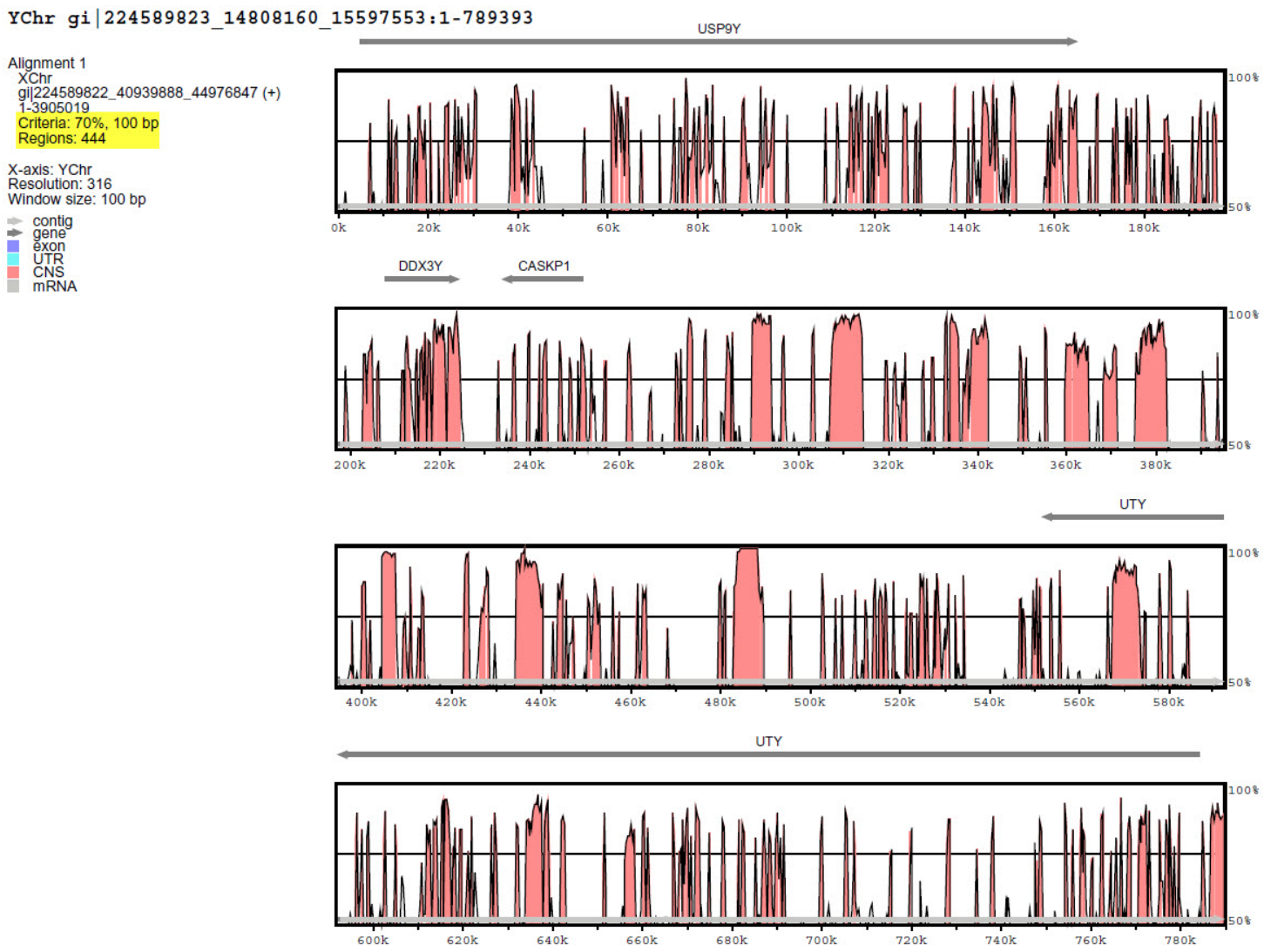


Abb. A-1: mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom I.
CNS-Kriterium: 100 bp lange Sequenz mit mind. 70% Homologie.

Human Y-chromosomale Sequenz aus GenBank Acc.No. „NC_000024.9“ (Pos. 14.808.160 - 15.597.553 = 789.393 bp). Human X-chromosomale Sequenz aus GenBank Acc.No. „NC_000023.10“ (Pos. 40.939.888 - 44.976.847 = 4.036.960 bp).

Genpositionen Y: *USP9Y* (+Ori): 14.813.160 - 14.972.764; *DDX3Y* (+Ori): 15.015.753 15.032.390; *CASKP1* (-Ori): 15.042.075 - 15.060.090; *UTY* (-Ori): 15.360.259 - 15.592.553.

Genpositionen X: *USP9X* (+Ori): 40.944.888 - 41.092.185; *DDX3X* (+Ori): 41.192.601 - 41.223.725; *NYX* (+Ori; nicht auf Y-Chromosom): 41.306.687 - 41.334.963; *CASK* (-Ori): 41.374.187 - 41.782.716; *UTX* (+Ori): 44.732.423 - 44.971.847. Ori = Orientierung.

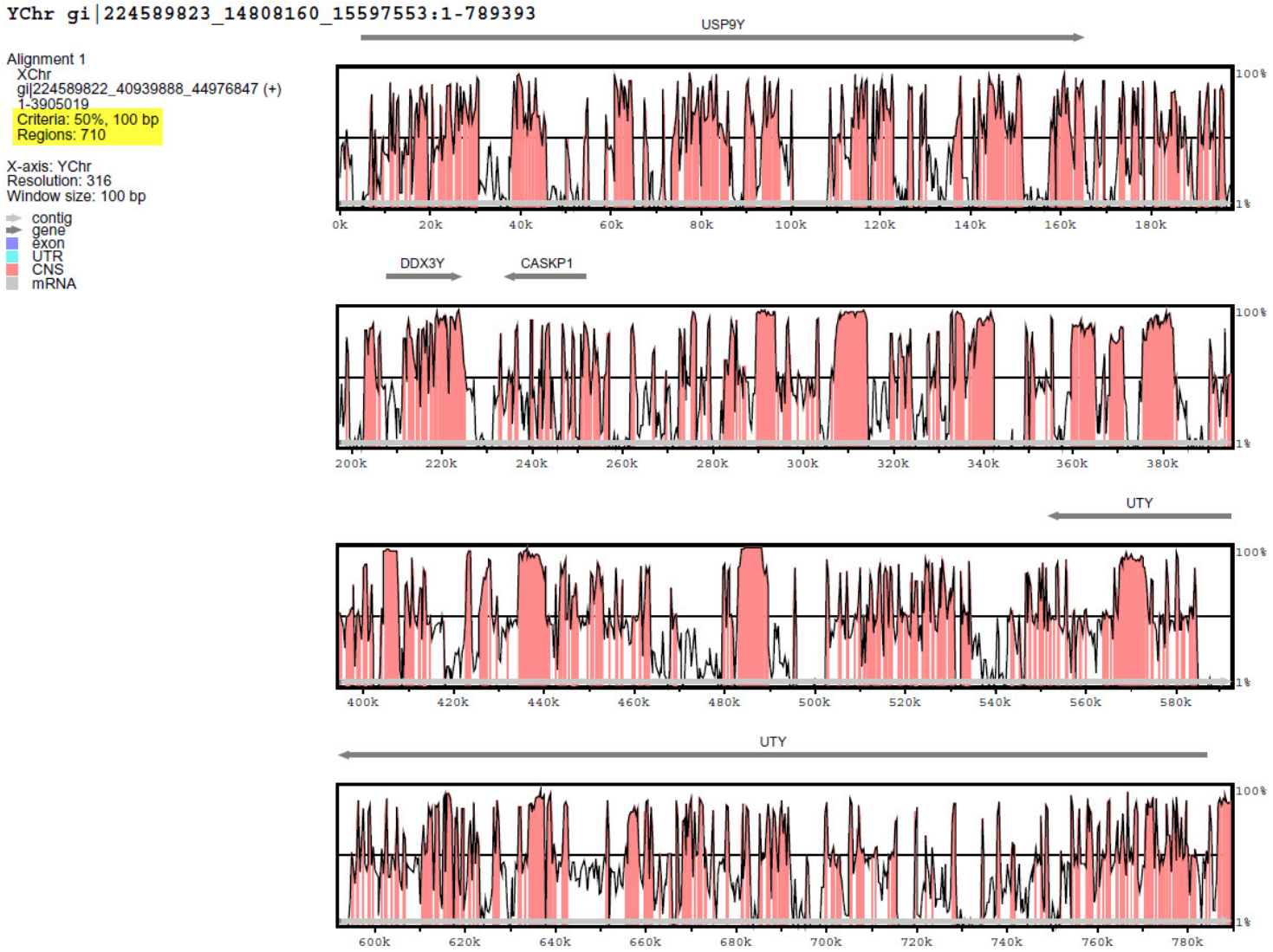


Abb. A-2: mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom II.

CNS-Kriterium: 100 bp lange Sequenz mit mind. 50% Homologie.

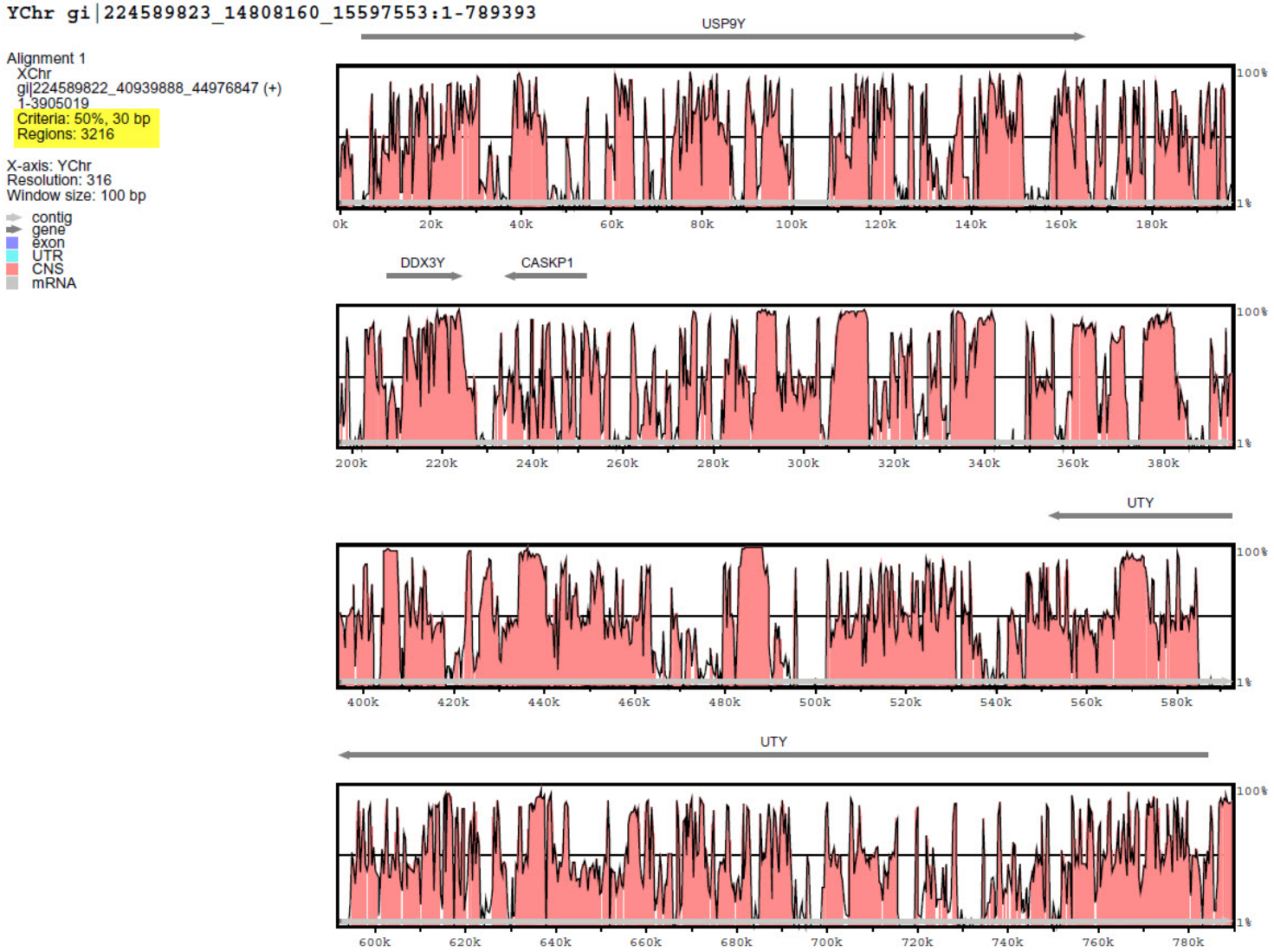
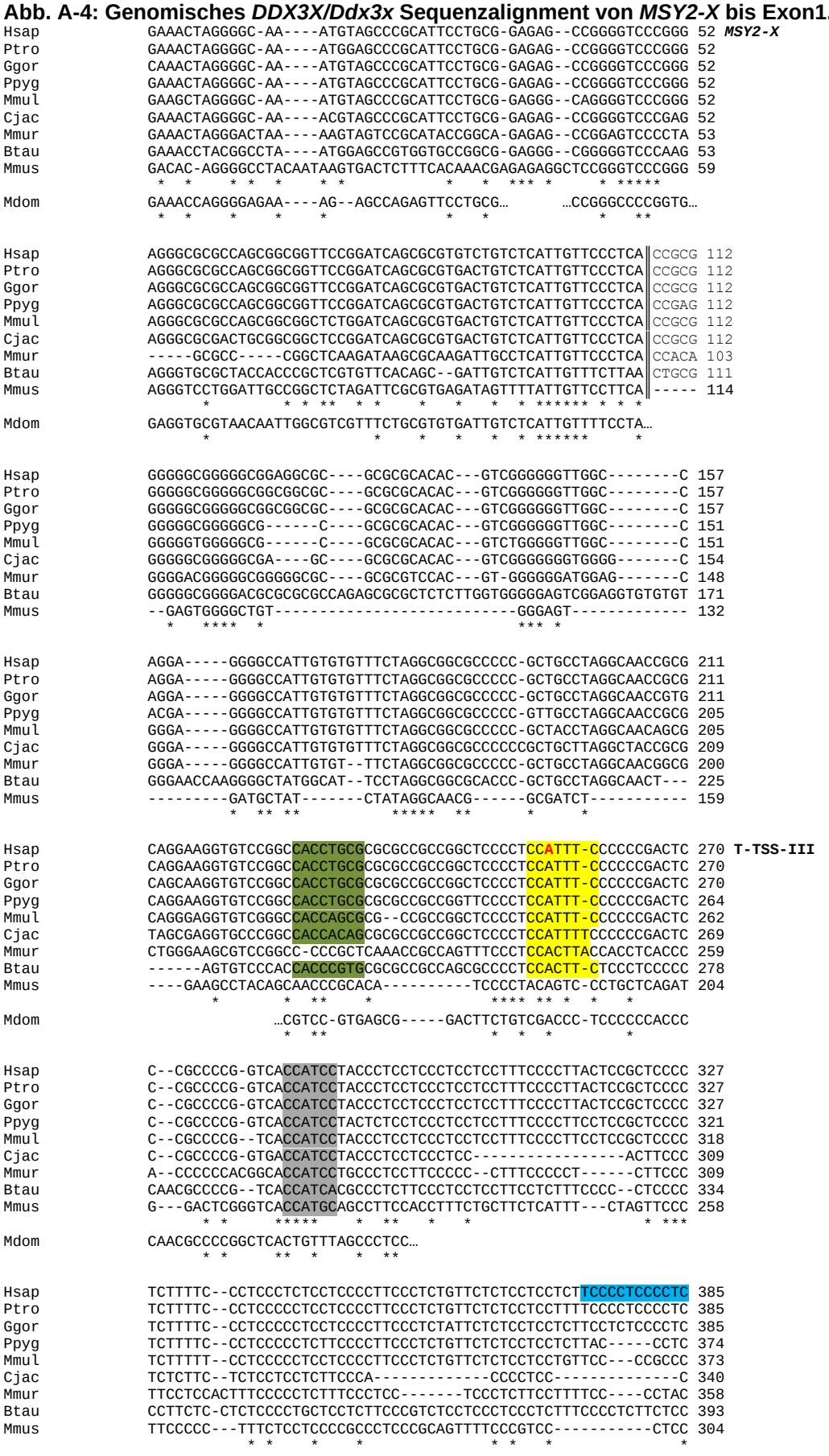


Abb. A-3: mVISTA Analyse Y-chromosomaler AZFa Locus [*USP9Y–DDX3Y–UTY*] im Vergleich zum homologen X-chromosomalen Gencluster [*USP9X–DDX3X–UTX*] im Humangenom III.

CNS-Kriterium: 30 bp lange Sequenz mit mind. 50% Homologie.

Abb. A-4: Genomisches *DDX3X/Ddx3x* Sequenzalignment von *MSY2-X* bis Exon1.



```
Hsap            CCCCG-TCCGGAGCACTCTATATTCAAGCCA-CCGTTTCCT--GCTTCACAAAATGG-CC 440 uORF1
Ptro            CCCCG-TCCGGAGCACTCTATATTCAAGCCA-CCGTTTCCT--GCTTCACAAAATGG-CC 440
Ggor            CCCCG-TCCCGAGCACTCTATATTCAAGCCA-CCGTTTCCT--GCTTCACAAAATGG-CC 440
Ppyg            CCCCG-TCCCGAGCACTCTGTATTCAATCCA-CCGTTTCCT--GCCTCACAAAATGG-CC 429
Mmul            CCCCG-TCCCGAGCACTCTAAATTCAAGCCA-CCGTTTCCT--GCTTCACAAAATGG-CC 428
Cjac            CCCTG-TCCCGAACACTCAATATTCACGCCA-CCGTTTCCT--GCTTCACAAAATGG-CC 395
Mmur            CCCAT-TTCCGATCACT----ATTCAAGCCAACCCTTTCCCTGGTCCCACAAAAAGGGCC 413
Btau            CCCCGATTCCGAGCACTCTATTTTAAAACCA-CGGTTTCCT--GCTCCACAAAATGG-CC 449
Mmus            CTCCCATTCCCAGCACT----AGTCAAGCGA---GTTTCCCT-TCTTCACAAGATGG-CC 355
                * *   * *  * ****      * *  * *    *****       ***** * ** **
Mdom                                                    …GCTTTGCAAAATGG-CC
                                                           *** * ** **

Hsap            ACCGCACGCGACACCTACGGTCACGTGGCCTGCCG-CCCTCTCAGTTTCGGGAAT----C 495
Ptro            ACCGCACGCGACACCTACGGTCACGTGGCCTGCCG-CCCTCTCAGTTTCGGGAAT----C 495
Ggor            ACCGCACGCGACACCTACGGTCACGTGGCCTGCCG-CCCTCTCGGTTTCGGGAAT----C 495
Ppyg            ACCGCACGCGACACCTACGGTCACGTGGCCTGCCG-CCCTCTCGGTTTCGGGACT----C 484
Mmul            ACCGCACGCGACACCTGCGGTCACGTGGCCTGCCG-CCCTCTCGGTTTCGGGACTGT--C 485
Cjac            ACCG--CGCGACACCTGCGGTCACGTGGCCTGCCG-CCCTCTAGGTTTTGGGACT----C 448
Mmur            ACCGCGCTCGCCACCTGCGATCACATGACCTACCG-CCCTCCCGTTTTAGGGCCTCTTTC 472
Btau            TCCGCATTTGGCACCTGCGGCCACGTGACCCCCTGTCCCCCTGGGCTTCGGGCCCCTCAT 509
Mmus            ACCTTGCTGGACACCCACGGTCACGTGATC--CCA-----------------------C 389
                 **      * ****  **  *** **  *  *
Mdom            GCGCCCGGCT---TCTGCGGTCACGTGACCCTCCATATTCTCCGCTTTCGGCTTTTTT-C
                 *            *  **  *** **  *  *

Hsap            TGCCTAGCTCCCACTAAGGGGAGGCTACCCGCGGAAGAGCGAGGGCAGATTAG----ACC 551
Ptro            TGCCTAGCTCCCACTAAGGGGAGGCTACCCGCGGAAGAGCGAGGGCAGATTAG----ACC 551
Ggor            TGCCTAGCTCCCACTAAGGGGAGGCTACCCGCGGAAGAGCGAGGGCAGATTAG----ACC 551
Ppyg            TGCCTAGCTCCGACTTCGGGGAGGCTACACGCGGAAGAGCGAGGGTAGATTAG----ACC 540
Mmul            TGCCTAGCTCCGACTTCGGGGAGGCTACACGCGGAAGACCGAGGGCAGATTAG----ACC 541
Cjac            TGCCTAGCTCCGACTTCGAGGACGCTACACGCAGAAGAGCGAGGGAAGATTAG----ACT 504
Mmur            TGCCTAGCTCTGATTTAGGGGAGGCTACACGCGTAAGGGCGAGGGCTGATTAG----ACC 528
Btau            TGCCAAACTTGGATCTCGGGGCTTCTGCCCGGGAAACTACGTGGGCTGATTAG----TTT 565
Mmus            TGCC--------------------CTCCACGC---AGAGCCAGGGCTCATTAGTTAGGCC 426
                ****                    ** * **    *   *  ***   *****
Mdom            TGGATAGCTC…
                **

Hsap            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGAACTGAGAGAGGAAGAAGAGTGAAG 607
Ptro            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGAACTGAGAGAGGAATTAGAGTGAAG 607
Ggor            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGAACTGAGAGAGGAAAAAGAGTGAAG 607
Ppyg            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGAACTGAGAGAGGAAGAAGAGTGAAG 596
Mmul            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGAACTGAGAGAGGA---AGAGTGAAG 594
Cjac            GGAGAAATCCCACCACATC---TCCAAGCCC-GGGGACTGAGAGAGGA---AGAGTGAAG 557
Mmur            GGAGAAATCCCACCATATC---TCCAAGCCC-GGGAACTGGGAGAGAAAAAAGAGAGAGG 584
Btau            GGAGAAATCGCACAACCTT---TCAAAGTCTAGGGGACTGAGAGAGAA---GGAAAGAGT 619
Mmus            AGACAAAGCCCACTGCATCATCTCCACGCCC-GGGGACGGAGAGGAA----GGGACGAG- 480
                 ** *** * ***    *    ** * * *  *** ** * ***        *   **

Hsap            GCCAGTGTTAGGAAAAAAAAA---ACAAAAACAAAAAAAA------------------CG 646
Ptro            GCCAGTGTTAGGAAAAAAAAA---A---AAACAAAAAAAA------------------CG 643
Ggor            GCCAGTGTTAGGAAAAAAAAA---A----AAAAAACAAAA------------------CG 642
Ppyg            GCCAGTGTTAGGAAAAAAAAA---ACAAAAACAAAAAAAA------------------CG 635
Mmul            GCCAGTGATGGAAAAAAAAAAGAAAAGAAAAGAAAAAAAAAAACCCAAGAAAAAAAACCCG 654
Cjac            GCCACTGTTAGACAAAAAAAA--------AAAAAAAAAAG------------------G 590
Mmur            AACAGCGCTAGAGAAACAG-----------------------------------------G 604
Btau            GCT--CCCTAGAAAGAAAG-----------------------------------------G 637
Mmus            -----TGCTAGAAA----------------------------------------------G 490
                        * *  *                                              *

Hsap            AAAAACGAAAGCTGAGTGCATAGAGTTGGAAAGGGGAGCGAATGCGTAAGGTTGGAAAGG 706 uORF2
Ptro            AAAAACGAAAGCTGAGTGCATAGAGTTGGAAAGGGGAGC--------------------- 685
Ggor            AAAAACGAAAGCTGAGTGCATAGAGTTGGAAAGGGGAGC--------------------- 684
Ppyg            AAAAACTAAAGCTGAGTGCATACAGTTGGAAAGGGGAGCAAATGAATAAGGTTGGAAAGG 695
Mmul            AAAAACTAAAGCTGAGTGCATAGGGTTGGAAAGGGGAGC--------------------- 696
Cjac            AAAAACGAAAGCTGAGTACTTAGGGTTGGAAAGGGGAGT--------------------- 632
Mmur            AAGAGCGAAAACTGGATGCATAGGGGAGGAAAGAGGAGC--------------------- 646
Btau            AACAGTGAAAACTGAGGACGTTGGGAAAAAAAGGGGA----------------------- 677
Mmus            AAGGACGAAAACTGGGTACACAGGGGAGGAGCAGCCAAT--------------------- 532
                **     *** ***    *     *    *      *

Hsap            GGGGCGAAGAGGCCTAGGTTAACATTTTCAGGCGTCTTAGCCGGTGGAAAGCGGGAGACG 766 T/B-TSS-II
Ptro            -----GAAGAGGCCTAGGTTAACATTTTCAGGCGTCTTAGCCGGTGGGAAAGCGGGAGACG 737
Ggor            -----GAAGAGGCCTAGGTTAACATTTTCAGGCGTCTTAGCCGGTGGAAAGCGGGAGACG 736
Ppyg            GTAGCGAAGAAGCCTAGGTTAACATTTTCAGGCCTCTTAGCCGGTGGAAAGCGGGAGACG 755
Mmul            -----GAAGAGGCCTAGGTTAACATTTCCAGGCCTTTTAGCTGGTGGTAAACGGGAAACG 748
Cjac            -----GAAGAGGCTTAGATTAACATTTTCAGGGCTCCTAGCTGGCGGAAAGCGGGAGACT 684
Mmur            -----GAAGAGGCCTACGTTAA-ATTTCACGGCCTTGTAGCTGGCGGAAAGCGGGAG--G 695
Btau            -----AAA-AGGTCTAGGTTA-TATTTTAAGGTCTCCTAGCTGGCGGAAAATGGGAGGGG 727
Mmus            -----CAA----CCTAGTTTAA-AGCTTAGGGCCTTAAGGCTAGAGGA---CGGTCAGCA 576
                      **      **  ***  *  *   **  *    **  * **     **
```

```
Hsap            -CAAGTTCTCGCGAGATCTCGAGAACTCCGAGGC-TGAGACTAGGGTTTTAGCGGAGAGC 824
Ptro            -CAAGTTCTCGCGAGATCTCGAGAACTCCGAGGC-TGAGACTAGGGTTTTAGCGGAGAGC 795
Ggor            -CAAGTTCTCGCGAGATCTCGAGAACTCCGAGGC-TGAGACTGGGGTTTTAGCGGAGAGC 794
Ppyg            -CAAGTTCTCGCGAGATCTCGAGAACTCCGAGGC-TGAGACTAGGGTTTTAGCGGAGAGC 813
Mmul            -CAAGTTCTCGCGAGATCTCGAGACCTCCGAGGC-TGAGACTACGGTTTTAGCGGAGAGC 806
Cjac            -CCAGTTCTCGCGAGATCTCGAGAACTCCGAGGC-TGAGAATAGGGTTTTAGCAGAGAGT 742
Mmur            -CAAGTTCTCGCGAGATCTCGAGAACTCTAAGAC-CGAGGCTAGGTTATTAGCGGAGAAC 753
Btau            ATGAGGTCTCGCGAGATCTCGAGACCTCCCAGGCTTGAGAA-AAGGTAATAGCGGAGGAC 786
Mmus            ATCATTTCTCGCGAGATCTCGAGAA----AACAG-TAAGATTAGGGGACCGGTGGAGACC 631
                   *  ******************      *      **     *      *  ***

Mdom                …TTTTCGCGAGATCTTGAGCACTCCGAGTC-T…
                      * **********  ***       *

Hsap            ACGGGAAGTGTAGCTCGAGAGAACTGGGACAGCATTTCGCACCCTAAGCTCCAAGGCAGG 884
Ptro            ACGGGAAGTGTAGCTCGAGAGAACTGGGACAGCATTTGGCACCCTAAGCTCCAAGGCAGG 855
Ggor            ACGGGAAGTGTAGCTCGAGAGAACTGGGACAGCATTTGGCACCCTAAGCTCCAAGGCAGG 854
Ppyg            ACGGGAAGTGTAGCTCGAGAGAACTGGGACACTATTTGGCACCCTAAGCTCCAAGGCAGG 873
Mmul            ACGGGAAGTGTAGCTCGGGAGAACTGGGACAGCATTTGGCGCTCCAAGCTCCAAGACAGG 866
Cjac            AGGGGAAGTGTAGCTCGAGATAACTGGGACAGCATTTGGCGCCCTGAGCTCCAAGGCAGG 802
Mmur            AGGAAAAGTGAAGTTCAGGCGAACTAGGACAGCATTTGGCGCT-TGAGCTTCAAGGCAGG 812
Btau            AGGGGAAGTTGCACTCAAGCGATCTAGGACAGCATTTAGCGC--TGAGTTCCAAGGTAGG 844
Mmus            AGGGGATAGCAAACCCAGGAGAACCAGAACCACAT---ATACCCAAGGCTTCTTTATGAG 688
                * *  *          *  *  * *  * **   **       *     * * *       *

Hsap            ACTGCTAGG----GG-CGACAGGACTAAGTAGGAAATCCCTTGAGCTTAGACCTGAGGGA 939
Ptro            ACTGCTAGG----GG-CGACAGGACTAAGTAGGAAATCCCTTGAGGTTAGACCTGAGGGA 910
Ggor            ACTGCTAGG----GG-CGACAGGACTAAGTAGGAAATCCCTTGAGGTGAGACCTGAGGGA 909
Ppyg            ACTGCTAGG----GG-CGACGGGACTAAGTAGGAAATCCCTTGTGGTTAGACCTGAGGGA 928
Mmul            ACTGCTAGG----GG-CGACGGGACTAAGTAGGAAATCCCGTGAGGTTAGACCTGAGGGA 921
Cjac            ACGACTAGT----GG-CGACGGGACTGAGTAGGAAATCCCGTGAGATTAGACCTGAGGGA 857
Mmur            ACGACGAGG----GGACGACAGGACTGACTAGGAAATCCCGCGAGGTTAG----GACAGA 864
Btau            ACGACGAGC----CGCAGACGCGACCGACTAGGAAATCCCGCGAGGTTAGGACTAAACGA 900
Mmus            CCGACGAGGACGCGGGTAACTGGACTGATTAGGAAATCTCACGGGACTCAGCC-GAGC-A 746
                 *  * **       *   **  ***  * ********* *  * *          *   *

Hsap            GCGCGCAGTAGCCGGGCAGAAGTCGCCGCGACAGGGA-ATTGCGGT-GTGAGAGGGAGGG 997
Ptro            GCGCGCAGTAGTCGGGCAGAAGTCGCCGCGACAGGGA-ATTGCGGC-GTGAGAGGGAGGG 968
Ggor            GCGCGCAGTAGCCGGGCAGAAGTCGCCGCGACAGGGA-ATTGCGGT-GTGAGAGGGAGGG 967
Ppyg            GCGCGCAGTAGCCGGGCAGAAGTCGCCGCGACAGGGA-ATTGCAGT-GTGAGAGGGAGGG 986
Mmul            GCGCGCAGTAGCCGGGCATTAGTCGCCGCGACAGGGG-ATTGCAGT-GTGAGAGGGAGGG 979
Cjac            GCGCGCAGTAACTGGCCAGAAGTCGCCGCGGTGGGGG-ATTGCGGT-GTGAAAGGGAGGG 915
Mmur            GCACGCAGTAGCTAGGCAGAAGTCGTAGTGGTAGGGGGAGTGCGGGAGTGAGAGGGAGGG 924
Btau            GCGCGCAGTAGCTAGGCAAAAGTCGTAGCGGCGGGAGGAGTGCGGGTGTGAGAGGGCGGG 960
Mmus            ACGCACCGTAAGCAGGCAGAGCTCCTAACGGCAGTGGCCGGCCGCCCGGGAGAGG-AGGG 805
                 * * * ***    * **    **     *   *        *    * ** ***  ***

Hsap            CACACGTTGTACGTGCTGACGTAGCCGGCTTTCCAGCGGGTATATTAG-ATCCGTGGCCGC 1057
Ptro            CACACGTTGTACGTGCTGACGTAGCCGGCTTTCCAGCGGGTATATTAG-CTCCGTGGCCGC 1028
Ggor            CACACGTTGTACGTGCTGACGTAGCCGGCTTTCCAGCGGGTATATTAG-ATCCGTGGCCGC 1027
Ppyg            CATACGTTGTACGTGCTGACGTAGCCGGCTTTCCAGCGGGTATATTAG-ATCCGTGGCCGC 1046
Mmul            CACACGTTGTACGTGCTGACGTAGCGGGCTTTCAAGCGGGTATATTAG-ATCCGTGGCCGC 1039
Cjac            CACACGCTGTACGTGCTGACGTAGCAGGGTTTCGAGCGGGTATATTAG-GTCCGTGGCCGC 975
Mmur            AGTACGTCTTACGTGCTGACGTAACAGTCTTTCGAGTGGGTATATTAG-ATCCGTGGCCGC 984
Btau            CACACGTCGTACGTGCTGACGTAGCAGTCACTCGAGCGGGTATATTAG-ATCCGTGGCCGC 1020
Mmus            CACACGTCTTACGTGCTGACGTAGCAGTCCTCCGAGTCGGTATATAAG-GTCGGTGGCCGC 865
                   ***   ************** * *     * **  ******* **   ** ********

Mdom                                             …GGTCATATTAAGGGCCGTGGCCGC
                                                   *  **** *    * ********

Hsap            GCGGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAAACAC- 1116 TSS-I
Ptro            GCGGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAAACAC- 1087
Ggor            GCGGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAAACAC- 1086
Ppyg            GCGGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAAACACA 1106
Mmul            GCGGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAGACACA 1099
Cjac            GCCGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGTCTTCGCGGTGGAACAAAAACA 1035
Mmur            GCCGTGCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGAAACAAACACA 1044
Btau            GCCGTGCATTTCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGAACCCAACACA 1080
Mmus            GCGGTGCGCTTCAAAGCCGCAGTTCTCCCGTGAGAGGGCCTTTGCGGTAGAGGCAGCCT- 924
                ** ****  * ** ************************ *** *****  *

Mdom            GCGCGGGNNNNNNNNNN
                ***  *

Hsap            -TCGCTTAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCA 1175
Ptro            -TCGCTTAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCA 1146
Ggor            -TCGCTTAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCA 1145
Ppyg            CTCGCTTAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCA 1166
Mmul            CTCGCTTAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCG 1159
Cjac            CTCGCTCAGCAGCAGAAGACACCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCG 1095
Mmur            CTCGCTCAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCG 1104
Btau            TTAGCTCAGCAGCGGAAGACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCG 1140
Mmus            -TTGCTCAGCAGCGGAAGACTCCGATTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCA 983
                 * *** ****** ****** **** ********************************
```

```
Hsap            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1205
Ptro            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1176
Ggor            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1175
Ppyg            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1196
Mmul            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1189
Cjac            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1125
Mmur            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1134
Btau            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1170
Mmus            GTGGAAAATGCGCTCGGGCTGGACCAGCAG 1013
                ******************************
```

Abb. A-4: Genomisches *DDX3X/Ddx3x* Sequenzalignment von *MSY2-X* bis Exon1.

Human *DDX3X* Sequenz aus BAC-Klon „RP13-13A3“ (Pos. 29.570 - 30.774 = 1.205 bp). Im Bereich der kartierten *DDX3X* Core-Promotermotive wurden genomische Abschnitte der Opossum (Mdom) *Ddx3x* Sequenz eingefügt. Mdom Sequenz aus Chromosom 4 Referenzsequenz NW_001581956.1 (Pos. 22,34 Mb - 22,33 Mb, rev-komplementär). N = Sequenzierungslücke. Gelb = Initiator-Motiv, dunkelgrün = TATA-Motiv, grau = DPE-Motiv, orange = BRE^{d}, blau = Human Pentanukleotidrepeat $(TCCCC)_3$, rot = Translationsstartcodon ATG, hellgrün = Translationsstopcodon. Human 5´UTR Splicesignal = pink und Fettdruck. *MSY2-X* Sequenz ist durch horizontalen Strich getrennt. Hsap = *Homo sapiens*; Ptro = *Pan troglodytes*; Ggor = *Gorilla gorilla*; Ppyg = *Pongo pygmaeus*; Mmul = *Macaca mulatta*; Cjac = *Callithrix jacchus*; Mmur = *Microcebus murinus*; Btau = *Bos taurus*; Mmus = *Mus musculus*; Mdom = *Monodelphis domestica*.

Abb. A-5: potentielle uATGs in *DDX3X* Promoterkonstrukt I (2,44 kb).

ATGACTGGCGAGAGCTACTGGCATTTAGTGGGCAGGAGCCAGGCACGCTAGGATGTTCTATGATGCTGGT
CAGTCCCACAGAGCGAATAATTGAATCACTGGACACATGTCGGTGAAAAACTAGTTCATTATCATCTAAG
CCTAGAACTAACACTGTTTTACATATGAACACAAAGTATCTTTGCCCAGTTTCCATATATCCTGAATTTT
CCACAAATGCAGGTACTGTTTCAATTGAAGGAAGATTGTACTGTTTTATTTAAAAATTATTAATAGTTCA
TCATTTTGGAAAATGAGGTTCCTAGTAGCAATGCTGCTCATGGAACTTGAGTTGTGAAATCAACACACCT
TTCCAGTTTGCATTTGTGGCTCTACTGGGCAAGTGTCTTCCCACCAGTTAGTGTAGCTGTGCGGCGGCAT
TTACATATTATAATACATTGTTTTATTATAATCATTTTGCTTTCGTTTGTTGACAGATCATGTTTTTTGA
AATTGTGTTTGTAGGTTTACATAAACAATTTAGGTTTTGTTATTTCAGTATAGGTAAAGGGGGCGTTAGA
AAATACTCGTTATGAAAAGGAGTTGGGTCTTTCACAGGGTTATGGAATGTAGCAGACTCCTTTTATAGAT
GAGCAAATTGAGACCCAAAGCCGTTAAACAACTTGCCCCCGACAGTGTGATATTCTCTCCCTATCTCGTT
CGAATTTCCGTCTCGATTTTAAAAATTGAGAAGGGAGAGAAGTTATGTGTATAGTAAATTATGTGATGAG
CACGAACTATTTGAATTCTGGCAGCAGAAGACCTTTTGTGGGGGTGGTGGAAACCACTGCGTTTTATTTA
ATTTCGTAGGGCTGAGGGTCAGAAGAGAGGGGAAAGCGGTCATAAGAAACTACGGATGTGCGGGCGCCGG
CTCAAAGCCGGGAATCCACTAGTAAACACCTTCTTCTATTTAAGGGCCTGCGGCACCGAAGCCGGGACTT
GGGGTGGCCTGCCTAGGAGGGAACCACAGGAAGAGGTTCCACCGACGGTAGAAACCCGGTTCCTCCTGTT
CCTCATCCACCGCAGCGGCCTGAGAGTCACCCCGAGCCCAGGCCGGTAACCCCGCTGCAGCCCCTCCGGC
AGGCCCAGGTTGCACCGGCTACAAACATTCCCGACACCCTTGGGCCCTAGAAGAGCCGTTCGCCACCCGC
GGGGTCCCCGCAACGCCGGTGGTCACCGGCCCTAAGCCGCTGGGCGCCCCGCGGGAGGAAACTAGGGGCA
AATGTAGCCCGCATTCCTGCGGAGAGCCGGGGTCCCGGGAGGGCGCGCCAGCGGCGGTTCCGGATCAGCG
CGTGTCTGTCTCATTGTTCCCTCACCGCGGGGGGCGGGGGCGGAGGCGCGCGCGCACACGTCGGGGGGTT
GGCCAGGAGGGGCCATTGTGTGTTTCTAGGCGGCGCCCCCGCTGCCTAGGCAACCGCGCAGGAAGGTGTC
CGGCCACCTGCGCGCGCCGCCGGCTCCCCTCC|ATTTCCCCCCGACTCCCGCCCCGGTCACCATCCTACC
CTCCTCCCTCCTCCTTTCCCCTTACTCCGCTCCCCTCTTTTCCCTCCCTCTCCTCCCCTTCCCTCTGTTC
TCTCCTCCTCTTCCCCTCCCCTCCCCCGTCCGGAGCACTCTATATTCAAGCCACCGTTTCCTGCTTCACA
AAATGGCCACCGCACGCGACACCTACGGTCACGTGGCCTGCCGCCCTCTCAGTTTCGGGAATCTGCCTAG
CTCCCACTAAGGGGAGGCTACCCGCGGAAGAGCGAGGGCAGATTAGACCGGAGAAATCCCACCACATCTC
CAAGCCCGGGAACTGAGAGAGGAAGAAGAGTGAAGGCCAGTGTTAGGAAAAAAAAAACAAAAACAAAAAA
AACGAAAAACGAAAGCTGAGTGCATAGAGTTGGAAAGGGGAGCGAATGCGTAAGGTTGGAAAGGGGGGCG
AAGAGGCCTAGGTTAACATTTTCAGGCGTCTTAGCCGGTGGAAAGCGGGAGACGCAAGTTCTCGCGAGAT
CTCGAGAACTCCGAGGCTGAGACTAGGGTTTTAGCGGAGAGCACGGGAAGTGTAGCTCGAGAGAACTGGG
ACAGCATTTCGCACCCTAAGCTCCAAGGCAGGACTGCTAGGGGCGACAGGACTAAGTAGGAAATCCCTTG
AGCTTAGACCTGAGGGAGCGCGCAGTAGCCGGGCAGAAGTCGCCGCGACAGGGAATTGCGGTGTGAGAGG
GAGGGCACACGTTGTACGTGCTGACGTAGCCGGCTTTCCAGCGGGTATATTAGATCCGTGGCCGCGCGGT
GCGCTCCAGAGCCGCAGTTCTCCCGTGAGAGGGCCTTCGCGGTGGAACAAACACTCGCTTAGCAGCGGAA
GACTCCGAGTTCTCGGTACTCTTCAGGGATGAGTCATGTGGCAGTGGAAAATGCGCTCGGGCT‖acta***GC***
TAGCCTCGAGGATATCAAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCGGTACTGTTGGTAAAGCCAC
CATGG…

Abb. A-5: potentielle uATGs in *DDX3X* Promoterkonstrukt I (2,44 kb).

DDX3X Promotersequenz, BAC-Klon RP13-13A3, Pos. 28.323 - 30.764. uATGs sind gelb markiert. Natives *DDX3X* ATG und *Luc2* ATG sind rot markiert. Horizontaler Strich markiert *DDX3X* T-TSS-III, horizontaler Doppelstrich markiert Ende der *DDX3X* Sequenz. Linker aus Primer 1829R ist unterstrichen und in Kleinbuchstaben. *NheI* Schnittstelle ist kursiv und in Fettdruck. Stromabwärts von *NheI* ist Sequenz aus pGL4.10 (GenBank Acc.No. AY738222.1), bis einschließlich *Luc2* ATG, gezeigt. Es sind 25 uATGs möglich.

Tab. A-1: Liste der Nukleotide in der Schreibweise des „Ambiguity Code“.

Dieser Code erlaubt die Angabe von Vieldeutigkeiten der einzelnen Nukleotidpositionen in einer DNA-Sequenz.

Nukleotidabkürzung	Bedeutung	Beschreibung
A	Adenin	
C	Cytosin	
G	Guanin	
T	Thymin	
S	strong	C oder G; drei H-Brücken
W	weak	A oder T; zwei H-Brücken
K	Ketogruppe	G oder T
M	Aminogruppe	A oder C
R	Purin	A oder G
Y	Pyrimidin	C oder T
D	nicht C	A, G oder T
V	nicht T	A, C oder G
N oder X	alle Nukleotide erlaubt	A, C, G oder T

Tab. A-2: Vergleich *DDX3Y* BAC-Positionen mit Y-Chromosom Referenzsequenz.

BAC 475I1	Y-Chr.NC_000024.9, GRCh37.p2 „primary reference assembly“
MSY2 = Beginn 53.854	15.015.556
MSY2 Ende = 54.262	15.015.964
Start *MSY2-3*	15.015.767
Start *MSY2-4*	15.015.866
T-TSS-II in *MSY2-2* = 54.051	15.015.753
T-TSS-I = 54.317	15.016.019
Exon-Ta Ende = 54.623	15.016.325
Exon-Tb Ende = 54.722	15.016.424
TSS-I-(T) = 54.911	15.016.613
TSS-Ia = 55.032	15.016.734
TSS-Id = 55.104	15.016.806
Exon1a Start = 55.067	15.016.769
Exon1b Start = 55.144	15.016.846
ATG = 55.146	15.016.848
Ende Exon1 = 55.190	15.016.892
Start Exon17 = 68.253	15.029.955
TGA Stopcodon = 68.330	15.030.032
PAS1 = 68.553	15.030.255
PAS2 = 68.718	15.030.420
PAS3 = 70.680	15.032.382
Ende Exon17 = 70.688	15.032.390

Tab. A-3: Vergleich *DDX3X* BAC-Positionen mit X-Chromosom Referenzsequenz.

BAC RP13-13A3	**X-Chr. NC_000023.10, GRCh37.p2 „primary reference assembly“**
Anfang *MSY2-X* = 29.570	41.192.346
Ende *MSY2-X* = 29.676	41.192.452
T-TSS-III = 29.825	41.192.601
T/B-TSS-II = 30.314	41.193.090
TSS-I = 30.639	41.193.415
ATG = 30.730	41.193.506
Ende Exon1 = 30.774	41.193.550
TGA Stopcodon = 44.194	41.206.970
PAS1 = 44.584	41.207.303
PAS 2/3 = 46.658	41.209.429
PAS4 = 46.728	41.209.503
Ende Exon17 = 46.747	41.209.523 (16.923 bp)
PAS5 = 713 (in BAC-Klon RP1-169I5; GenBank Acc.No. Z93015.9)	41.223.699
Ende Exon 17c = 739	41.223.725 (31.125 bp)

Tab. A-4: Human *DDX3Y* Exon-Intronstruktur in BAC-Klon „475I1“ (GenBank Acc.No. AC004474.1).
Als 5´Grenze von *DDX3Y* Exon-T wurde die Position von T-TSS-II in *MSY2-2* angegeben. *DDX3Y* Exon-T Positionsangaben sind für Exon-Ta-1b Splicevariante angegeben. Als eindeutige 5´Grenze von Exon1 wurde die *DDX3Y* TSS-I-(T) Position angegeben. Human DDX3Y Protein hat 660 AS, im kodierenden RNA-Bereich besteht eine 88%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		**Länge in bp**		**Homologie zu Human *DDX3X***
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: 54.051 - 54.623		573		---
	Intron-T		520	
1: 54.911 - 55.190		280		77% mit *DDX3Y* TSS-Ib Transkript
	1		2.555	
2: 57.746 - 57.803		58		81%
	2		1.765	
3: 59.569 - 59.616		48		93%
	3		2.432	
4: 62.049 - 62.178		130		86%
	4		758	
5: 62.937 - 63.092		156		87%
	5		80	
6: 63.173 - 63.272		100		86%
	6		655	
7: 63.928 - 64.063		136		88%
	7		710	
8: 64.774 - 64.859		86		84%
	8		234	
9: 65.094 - 65.192		99		92%
	9		84	
10: 65.277 - 65.437		161		95%
	10		402	
11: 65.840 - 65.984		145		89%
	11		108	
12: 66.093 - 66.237		145		88%
	12		233	
13: 66.471 - 66.652		182		87%
	13		74	
14: 66.727 - 66.844		118		92%
	14		272	
15: 67.117 - 67.270		154		90%
	15		342	
16: 67.613 - 67.752		140		85%
	16		500	
17: 68.253 - 70.688 (≙ PAS3)		2.436		87% mit *DDX3X* Exon17 PAS2/3
TGA: 68.330				

Tab. A-5: Schimpanse *DDX3Y* Exon-Intronstruktur in BAC-Klon „CH251-128L22“ (GenBank Acc.No. AC146254.2) revers-komplementär.

Als 5´Grenze von *DDX3Y* Exon-T wurde die Position von Primer P2598 in *MSY2-1* angegeben. Als eindeutige 5´Grenze von Exon1 wurde die *DDX3Y* TSS-I-(T) Position angegeben. Schimpanse DDX3Y Protein hat 660 AS, im kodierenden RNA-Bereich besteht eine 99%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: 60.095 - 59.546		550		95%
	Intron-T		519	
1: 59.258 - 58.980		279		96%
	1		2.574	
2: 56.405 - 56.348		58		98%
	2		1.766	
3: 54.581 - 54.534		48		100%
	3		2.457	
4: 52.076 - 51.947		130		99%
	4		758	
5: 51.188 - 51.033		156		98%
	5		80	
6: 50.952 - 50.853		100		100%
	6		653	
7: 50.199 - 50.064		136		99%
	7		710	
8: 49.353 - 49.268		86		99%
	8		272	
9: 48.995 - 48.897		99		99%
	9		84	
10: 48.812 - 48.652		161		100%
	10		383	
11: 48.268 - 48.124		145		100%
	11		108	
12: 48.015 - 47.871		145		99%
	12		233	
13: 47.637 - 47.456		182		98%
	13		72	
14: 47.383 - 47.266		118		100%
	14		276	
15: 46.989 - 46.836		154		99%
	15		342	
16: 46.493 - 46.354		140		99%
	16		491	
17: 45.862 - 43.428 (≙ PAS3 Human)		2.435		98%
TGA: 45.785				

Tab. A-6: *Macaca DDX3Y* Exon-Intronstruktur in BAC-Klon „CH250-541A11“ (GenBank Acc.No. AC213321.3) revers-komplementär.

Als 5´Grenze von *DDX3Y* Exon-T wurde die Position von Primer P1329 in *MSY2-2* angegeben. Als eindeutige 5´Grenze von Exon1 wurde die Position von Primer P2389 angegeben. *Macaca* DDX3Y Protein hat 661 AS, im kodierenden RNA-Bereich besteht eine 98%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: 83.984 - 83.520		465		93% mit Human T-TSS-II Transkript
	Intron-T		525	
1: 83.252 - 82.948		305		90%
	1		2.535	
2: 80.412 - 80.352		61		95%
	2		1.908	
3: 78.443 - 78.396		48		100%
	3		2.168	
4: 76.227 - 76.098		130		98%
	4		763	
5: 75.334 - 75.179		156		96%
	5		80	
6: 75.098 - 74.999		100		96%
	6		630	
7: 74.368 - 74.233		136		100%
	7		707	
8: 73.525 - 73.440		86		97%
	8		223	
9: 73.216 - 73.118		99		98%
	9		84	
10: 73.033 - 72.873		161		100%
	10		464	
11: 72.408 - 72.264		145		100%
	11		107	
12: 72.156 - 72.012		145		97%
	12		244	
13: 71.767 - 71.586		182		97%
	13		82	
14: 71.503 - 71.386		118		99%
	14		575	
15: 70.810 - 70.657		154		99%
	15		340	
16: 70.316 - 70.177		140		96%
	16		503	
17: 69.673 - 67.230 (≙ PAS3 Human)		2.444		95%
TGA: 69.596				

Tab. A-7: *Callithrix DDX3Y* Exon-Intronstruktur in BAC-Klon „CH259-161K20“ (GenBank Acc.No. AC225609.4) revers-komplementär.

Als 5´Grenze von *DDX3Y* Exon-T wurde die Position von T-TSS-I angegeben. Als 5´Grenze von *DDX3Y* Exon1 wurde die Position von P2403. *Callithrix* DDX3Y Protein hat 654 AS, im kodierenden RNA-Bereich besteht eine 94%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: 52.405 - 52.161		245		79% mit Human T-TSS-I Transkript
	Intron-T		508	
1: 51.938 - 51.606		333		75% mit Human TSS-I-(T) Transkript
	1		2.500	
2: 49.105 - 49.045		61		93%
	2		986	
3: 48.058 - 48.011		48		94%
	3		2.206	
4: 45.804 - 45.690		115		93%
	4		1.280	
5: 44.409 - 44.254		156		92%
	5		80	
6: 44.173 - 44.074		100		95%
	6		637	
7: 43.436 - 43.301		136		96%
	7		732	
8: 42.568 - 42.483		86		93%
	8		229	
9: 42.253 - 42.155		99		93%
	9		84	
10: 42.070 - 41.910		161		98%
	10		405	
11: 41.504 - 41.360		145		97%
	11		104	
12: 41.255 - 41.111		145		94%
	12		235	
13: 40.875 - 40.694		182		93%
	13		73	
14: 40.620 - 40.503		118		97%
	14		285	
15: 40.217 - 40.064		154		97%
	15		358	
16: 39.705 - 39.572		134		88%
	16		214	
17: 39.357 - 36.907 (≙ PAS3 Human)		2451		88%
TGA: 39.280				

Tab. A-8: Rind *Ddx3y* Exon-Intronstruktur in BAC-Klon „CH240-56A13“ (GenBank Acc.No. AC217167.7) revers-komplementär.

Als 5´Grenze von Exon-T wurde die Position von Rind *Ddx3y* Primer P2590 angegeben. Als eindeutige 5´Grenze von Exon1 wurde die Position von „Rind *Ddx3y* TSS-I-(T)“ angegeben. Rind Ddx3y Protein hat 660 AS, im kodierenden RNA-Bereich besteht eine 89%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: 140.615 – 140.282		334		60% mit Human T-TSS-I Transkript
	Intron-T		398	
1: 139.971 – 139.747		225		67% mit Human TSS-I-(T) Transkript
	1		2.109	
2: 137.637 – 137.577		61		84%
	2		1.546	
3: 136.030 – 135.983		48		94%
	3		498	
4: 135.484 – 135.355		130		85%
	4		644	
5: 134.710 – 134.552		159		89%
	5		74	
6:134.477 – 134.378		100		92%
	6		406	
7:133.971 – 133.836		136		87%
	7		975	
8:132.860 – 132.775		86		91%
	8		233	
9:132.541 – 132.443		99		93%
	9		84	
10:132.358 – 132.198		161		89%
	10		387	
11:131.810 – 131.666		145		94%
	11		100	
12:131.565 – 131.421		145		88%
	12		228	
13:131.192 – 131.011		182		90%
	13		104	
14:130.906 – 130.789		118		93%
	14		165	
15:130.623 – 130.470		154		86%
	15		293	
16:130.176 – 130.043		134		83%
	16		221	
17:129.821 – 127.434 (≙ PAS3 Human)		2.388		80%
TGA: 129.744				

Tab. A-9: Maus *Ddx3y* Exon-Intronstruktur in BAC-Klon „RP24-208N6“ (GenBank Acc.No. AC145393.4) revers-komplementär.

Als 5´Grenze von Exon1 wurde die Position von *Ddx3y* Primer P2371 angegeben. Maus Ddx3y Protein hat 658 AS, im kodierenden RNA-Bereich besteht eine 88%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
Exon-T: ---				
	Intron-T: ---			
1: 93.858 – 93.700		159		60%
	1		2.096	
2: 91.603 – 91.543		61		77%
	2		1.373	
3: 90.169 – 90.122		48		98%
	3		2.461	
4: 87.660 – 87.531		130		83%
	4		731	
5: 86.799 – 86.644		156		87%
	5		426	
6: 86.217 – 86.118		100		90%
	6		9.056	
7: 77.061 – 76.926		136		92%
	7		2.516	
8: 74.409 – 74.324		86		88%
	8		409	
9: 73.914 – 73.816		99		93%
	9		75	
10: 73.740 – 73.580		161		94%
	10		385	
11: 73.194 – 73.050		145		93%
	11		109	
12: 72.940 – 72.796		145		89%
	12		365	
13: 72.430 – 72.249		182		87%
	13		69	
14: 72.179 – 72.062		118		94%
	14		891	
15: 71.170 – 71.017		154		86%
	15		95	
16: 70.921 – 70.791		131		80%
	16		202	
17: 70.588 – 67.955 (≙ PAS3 Human)		2.634		74%
TAG: 70.511				

Tab. A-10: Human *DDX3X* Exon-Intronstruktur in BAC-Klon „RP13-13A3“ (GenBank Acc.No. AL391647.16).

Als 5´Grenze von *DDX3X* Exon1 wurde die Position von T-TSS-III angegeben. Human DDX3X Protein hat 662 AS, im kodierenden RNA-Bereich besteht eine 88%ige Homologie zum Human *DDX3Y* Transkript.

BAC-Position		Länge in bp		Homologie zu Human *DDX3Y*
Exon	**Intron**	**Exon**	**Intron**	**Exon**
1: 29.825 - 30.774		950		77% mit *DDX3Y* TSS-Ib Transkript
	1		3.110	
2: 33.885 - 33.942		58		81%
	2		1.570	
3: 35.513 - 35.560		48		93%
	3		2.400	
4: 37.961 - 38.093		133		86%
	4		878	
5: 38.972 - 39.130		159		87%
	5		83	
6: 39.214 - 39.313		100		86%
	6		379	
7: 39.693 - 39.828		136		88%
	7		385	
8: 40.214 - 40.299		86		84%
	8		207	
9: 40.507 - 40.605		99		92%
	9		110	
10: 40.716 - 40.876		161		95%
	10		780	
11: 41.657 - 41.801		145		89%
	11		79	
12: 41.881 - 42.025		145		88%
	12		680	
13: 42.706 - 42.887		182		87%
	13		94	
14: 42.982 - 43.099		118		92%
	14		236	
15: 43.336 - 43.489		154		90%
	15		299	
16: 43.789 - 43.928		140		85%
	16		188	
17: 44.117 - 46.747 (≙ PAS4 Human)		2.631		87% mit *DDX3X* Exon17 PAS2/3
TGA: 44.194				

Tab. A-11: Schimpanse *DDX3X* Exon-Intronstruktur in BAC-Klon „CH251-3A20“ (GenBank Acc.No. AC183643.2).

Als 5´Grenze von Exon1 wurde die Position von *DDX3X* T-TSS-III angegeben. Schimpanse DDX3X Protein hat 662 AS, im kodierenden RNA-Bereich besteht eine 99,6%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
1: 720 - 1.640		921		96%
	1		3.105	
2: 4.746 - 4.803		58		100%
	2		1.587	
3: 6.391 - 6.438		48		100%
	3		2.433	
4: 8.872 - 9.004		133		100%
	4		878	
5: 9.883 - 10.040		159		98%
	5		83	
6: 10.124 - 10.223		100		100%
	6		379	
7: 10.603 - 10.738		136		100%
	7		385	
8: 11.124 - 11.209		86		100%
	8		207	
9: 11.417 - 11.515		99		100%
	9		110	
10:11.626 - 11.786		161		100%
	10		781	
11: 12.568 - 12.712		145		100%
	11		79	
12: 12.792 - 12.936		145		100%
	12		675	
13: 13.612 - 13.793		182		99%
	13		94	
14: 13.888 - 14.005		118		99%
	14		236	
15: 14.242 - 14.395		154		99%
	15		299	
16: 14.695 - 14.834		140		99%
	16		188	
17: 15.023 - 17.635 (≙ Human PAS4)		2.613		99%
TGA: 15.100				

Tab. A-12: *Macaca DDX3X* Exon-Intronstruktur in Ensembl *Macaca* X-Chromosom (Ensembl Assembly MMUL 1.0, Datenbankversion 62.10p ≙ GenBank Acc.No. NC_007878.1).

Als 5´Grenze von Exon1 wurde die Position von *DDX3X* T-TSS-III angegeben. *Macaca* DDX3X Protein hat 662 AS, im kodierenden RNA-Bereich besteht eine 99%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	Intron	Exon	Intron	Exon
1: 39.101.250 - 39.102.191		942		91%
	1		3.370	
2: 39.105.563 - 39.105.620		58		100%
	2		1.675	
3: 39.107.296 - 39.107.343		48		100%
	3		2.431	
4: 39.109.775 - 39.109.907		133		99%
	4		879	
5:39.110.787 - 39.110.945		159		99%
	5		78	
6:39.111.024 - 39.111.123		100		99%
	6		376	
7: 39.111.500 - 39.111.635		136		99%
	7		384	
8: 39.112.020 - 39.112.105		86		99%
	8		210	
9: 39.112.316 - 39.112.414		99		100%
	9		110	
10: 39.112.525 - 39.112.685		161		100%
	10		751	
11: 39.113.437 - 39.113.581		145		99%
	11		389	
12: 39.113.971 - 39.114.115		145		100%
	12		549 + Sequenzlücke	
13: 39.115.199 - 39.115.380		182		98%
	13		93	
14: 39.115.474 - 39.115.591		118		97%
	14		236	
15: 39.115.828 - 39.115.981		154		99%
	15		302	
16: 39.116.284 - 39.116.423		140		98%
	16		44 + Sequenzlücke	
17: 39.116.858 - 39.118.889		2.032 + Sequenzlücke		99% im vorhandenen Bereich
TGA: Sequenzlücke				

Tab. A-13: *Callithrix DDX3X* Exon-Intronstruktur in BAC-Klon „CH259-42H12“ (GenBank Acc.No. AC193908.1) revers-komplementär.

Als 5´Grenze von Exon1 wurde die Position von *DDX3X* T-TSS-III angegeben. *Callithrix* DDX3X Protein hat 662 AS, im kodierenden RNA-Bereich besteht eine 98%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	**Intron**	**Exon**	**Intron**	**Exon**
1: 97.588 - 96.717		872		90%
	1		3.151	
2: 93.565 - 93.508		58		100%
	2		1.388	
3: 92.119 - 92.072		48		100%
	3		2.134	
4: 89.937 - 89.805		133		98%
	4		887	
5: 88.917 - 88.759		159		99%
	5		83	
6: 88.675 - 88.576		100		99%
	6		376	
7: 88.199 - 88.064		136		99%
	7		655	
8: 87.408 - 87.323		86		99%
	8		128	
9: 87.194 - 87.096		99		100%
	9		123	
10: 86.972 - 86.812		161		98%
	10		655	
11: 86.156 - 86.012		145		99%
	11		82	
12: 85.929 - 85.785		145		98%
	12		597	
13: 85.187 - 85.006		182		96%
	13		85	
14: 84.920 - 84.803		118		98%
	14		241	
15: 84.561 - 84.408		154		97%
	15		293	
16: 84.114 - 83.975		140		97%
	16		181	
17: 83.793 - 81.172 (≙ Human PAS4)		2.622		98%
TGA: 83.716				

Tab. A-14: *Rind DDX3X* Exon-Intronstruktur in genomischen Contig „004.81.scaffold1“ (GenBank Acc.No. NW_001501904.1) revers-komplementär.

Als eindeutige 5´Grenze des hodenspezifisch verlängerten Exon1 wurde die Position von „Btau *Ddx3x* T-TSS-II“ angegeben. Rind Ddx3x Protein hat 661 AS, im kodierenden RNA-Bereich besteht eine 95%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	Intron	Exon	Intron	Exon
1: 82.883 - 82.419		465		83%
	1		2.774	
2: 79.644 - 79.587		58		100%
	2		1.533	
3:78.053 - 78.006		48		96%
	3		1.684	
4:76.321 - 76.189		133		94%
	4		958	
5: 75.230 - 75.075		156		92%
	5		80	
6: 74.994 - 74.895		100		95%
	6		639	
7: 74.255 - 74.120		136		97%
	7		1.204	
8: 72.915 - 72.830		86		94%
	8		172	
9: 72.657 - 72.559		99		92%
	9		108	
10: 72.450 - 72.290		161		94%
	10		314	
11: 71.975 - 71.831		145		93%
	11		80	
12: 71.750 - 71.606		145		92%
	12		181	
13: 71.424 - 71.243		182		96%
	13		92	
14: 71.150 - 71.033		118		96%
	14		277	
15: 70.755 - 70.602		154		94%
	15		239	
16: 70.362 - 70.223		140		91%
	16		192	
17: 70.030 - 67.414 (≙ PAS4 Human)		2.617		97%
TGA: 69.953				

Tab. A-15: Maus *DDX3X* Exon-Intronstruktur in BAC-Klon „RP23-158L19“ (GenBank Acc.No. AL833805.11).

Als eindeutige 5´Grenze von Exon1 wurde die Position von Maus *Ddx3x*TSS-I angegeben. Maus Ddx3x Protein hat 662 AS, im kodierenden RNA-Bereich besteht eine 93%ige Homologie zum Human *DDX3X* Transkript.

BAC-Position		Länge in bp		Homologie zu Human
Exon	Intron	Exon	Intron	Exon
1: 90.448 - 90.583		136		91% mit Human TSS-I Transkript
	1		3.020	
2: 93.604 - 93.661		58		98%
	2		679	
3: 94.341 - 94.388		48		92%
	3		1.071	
4: 95.460 - 95.592		133		91%
	4		1.058	
5:96.651 - 96.809		159		91%
	5		80	
6:96.890 - 96.989		100		94%
	6		221	
7:97.211 - 97.346		136		97%
	7		358	
8:97.705 - 97.790		86		90%
	8		208	
9:97.999 - 98.097		99		88%
	9		108	
10:98.206 - 98.366		161		93%
	10		454	
11:98.821 - 98.965		145		94%
	11		297	
12:99.263 - 99.407		145		88%
	12		197	
13:99.605 - 99.786		182		94%
	13		85	
14:99.872 - 99.989		118		89%
	14		337	
15:100.327 - 100.480		154		91%
	15		108	
16:100.589 - 100.728		140		90%
	16		110	
17:100.839 - 103.461 (≙ PAS4 Human)		2.623		96%
TGA:100.916				

Tab. A-16: Liste der potentiellen TF-Bindestellen in Human *DDX3Y* Promotersequenz.

BAC-Klon 475I1 (Position 52.646 - 55.190 = 2.545). Siehe beigefügte Daten-CD.

Tab. A-17: Liste der potentiellen TF-Bindestellen in Human *DDX3X* Promotersequenz.

BAC-Klon RP13-13A3 (Position 28.230 - 30.774 = 2.545). Siehe beigefügte Daten-CD.

Tab. A-18: Liste der potentiellen TF-Bindestellen in Schimpanse *DDX3Y* Promotersequenz.

Siehe beigefügte Daten-CD.

Tab. A-19: Liste der potentiellen TF-Bindestellen in *Macaca DDX3Y* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-20: Liste der potentiellen TF-Bindestellen in *Callithrix DDX3Y* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-21: Liste der potentiellen TF-Bindestellen in Rind *Ddx3y* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-22: Liste der potentiellen TF-Bindestellen in Maus *Ddx3y* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-23: Liste der potentiellen TF-Bindestellen in Schimpanse *DDX3X* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-24: Liste der potentiellen TF-Bindestellen in *Macaca DDX3X* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-25: Liste der potentiellen TF-Bindestellen in *Callithrix DDX3X* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-26: Liste der potentiellen TF-Bindestellen in Rind *Ddx3x* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-27: Liste der potentiellen TF-Bindestellen in Maus *Ddx3x* Promotersequenz.
Siehe beigefügte Daten-CD.

Tab. A-28: Liste der potentiellen GAGA TF-Bindestellen 3´ zu Human *DDX3Y* Genlocus in pot. S/MAR-Element Nr.8 (Abb. 4-55).
Human Y-Chromosom-Sequenz „NC_000024.9“ (Pos. 15.047.537 - 15.047.737 = 201 bp). TFBS bestimmt mit Genomatix MatInspector.

Matrix	Opt. threshold	Strand	Matrix sim.	Core sim.	Sequence
GAGA.01	0.78	-	0.845	1.0	gagagAGAGaaagaaaggaaggatg
GAGA.01	0.78	-	0.917	1.0	gagagAGAGagagaaagaaaggaag
GAGA.01	0.78	-	0.836	1.0	acgagAGAGagagagaaagaaagga
GAGA.01	0.78	-	0.924	1.0	gaacgAGAGagagagagaaagaaag
GAGA.01	0.78	-	0.792	0.75	aagaaCGAGagagagagagaaagaa
GAGA.01	0.78	-	0.827	0.75	aagaaAGAAcgagagagagagagaa
GAGA.01	0.78	-	0.797	0.75	gaaagAAAGaacgagagagagagag
GAGA.01	0.78	-	0.806	0.75	gagagAAAGaaagaacgagagagag
GAGA.01	0.78	-	0.847	1.0	taaagAGAGaaagaaagaacgagag
CDX2.02	0.85	+	0.872	1.0	ttctctctTTATttctttt
IRF3.01	0.85	-	0.937	1.0	gagagaaaaaGAAAtaaagag
NMP4.01	0.97	-	0.97	1.0	agAAAAagaaa
GAGA.01	0.78	-	0.79	1.0	gggagAGAGagaaaaagaaataaag
GAGA.01	0.78	-	0.809	1.0	aagggAGAGagagaaaaagaaataa
E2F.01	0.75	-	0.765	1.0	ggagagagaGAAAaaga
GAGA.01	0.78	-	0.807	0.75	gaaggAAAGggagagagagaaaaag
IRF3.01	0.85	-	0.851	0.758	gaaggaaaggGAGAgagagaa

PRDM5.01	0.71	-	0.755	1.0	gagggagggAAGGaaggaaagggagagag
PEA3.01	0.94	-	0.96	1.0	ggagggaAGGAaggaaaggga
GKLF.01	0.86	-	0.896	0.817	agggaaggaaggaAAGG
GAGA.01	0.78	-	0.794	0.75	gagggAGGGagggaaggaaggaaag
SPZ1.01	0.94	-	0.952	1.0	gGGAGggaagg
MAZ.01	0.9	-	0.914	1.0	gaggGAGGgaagg
CKROX.01	0.88	-	0.922	1.0	gagggaGGGAgggaagg
GAGA.01	0.78	-	0.796	0.75	gaaagAGGGagggagggaaggaagg
GAGA.01	0.78	-	0.834	1.0	gagaaAGAGggagggagggaaggaa
CKROX.01	0.88	-	0.957	1.0	gaaagaGGGAgggaggg
GAGA.01	0.78	-	0.848	0.75	gagagAAAGagggagggagggaagg
GAGA.01	0.78	-	0.848	0.75	aagagAGAAagagggagggagggaa
BARBIE.01	0.88	-	0.912	1.0	agagAAAGagggagg
GAGA.01	0.78	-	0.887	1.0	aaaagAGAGaaagagggagggaggg
GKLF.01	0.86	-	0.863	1.0	gagagaaagagggAGGG
BLIMP1.01	0.81	-	0.836	1.0	aagagaGAAAgagggaggg
GAGA.01	0.78	-	0.801	1.0	gaaaaAGAGagaaagagggagggag
GAGA.01	0.78	-	0.807	0.75	aagaaAAAGagagaaagagggaggg
NMP4.01	0.97	-	0.97	1.0	agAAAAagaga
GAGA.01	0.78	-	0.813	1.0	gaaagAGAGaaagaaagaaaaagag
GAGA.01	0.78	-	0.848	0.75	gacagAAAGagagaaagaaagaaaa
GAGA.01	0.78	-	0.808	0.75	aagagACAGaaagagagaaagaaag
GAGA.01	0.78	-	0.826	1.0	aaaaaAGAGacagaaagagagaaag
NFAT5.01	0.83	-	0.881	1.0	aaaGGAAaaaaagagacag
NMP4.01	0.97	-	0.978	1.0	ggAAAAaaaga
NFAT.01	0.95	-	0.966	1.0	aaaaaGGAAaaaaagagac
ELF3.01	0.91	-	0.923	1.0	agaaaaaaGGAAaaaaagaga
GAGA.01	0.78	-	0.795	0.75	gagagAAAGaaagaaagaaagaaaa
GAGA.01	0.78	-	0.858	1.0	gaaggAGAGaaagaaagaaagaaag
GAGA.01	0.78	-	0.798	0.75	gaaagAAAGaaagaaggagagaaag
GATA2.02	0.9	-	0.918	1.0	gaaaGATAgaaag
GAGA.01	0.78	-	0.78	0.75	gaaagAAAGaaagaaagatagaaag

Tab. A-29: Rohdaten *DDX3X* Promoter-Luziferaseassay (siehe Tab. 4-28).
Siehe beigefügte Daten-CD.

Tab. A-30: Rohdaten *DDX3X* Promoter-Luziferaseassay mit partiellen Promoterkonstrukten (siehe Tab. 4-30).
Siehe beigefügte Daten-CD.

Tab. A-31: Rohdaten *DDX3Y MSY2*-Luziferaseassay (siehe Tab. 4-34).
Siehe beigefügte Daten-CD.

Tab. A-32: Rohdaten *DDX3Y MSY2*-Mutations-Luziferaseassay (siehe Tab. 4-36).
Siehe beigefügte Daten-CD.

Tab. A-33: Rohdaten *DDX3X MSY2*-X Luziferaseassay (siehe Tab. 4-38).

Siehe beigefügte Daten-CD.

Tab. A-34: Rohdaten *DDX3X MSY2*-X Mutations-Luziferaseassay (siehe Tab. 4-40).
Siehe beigefügte Daten-CD.

Tab. A-35: Rohdaten *DDX3X MSY2*-X Mutations-Luziferaseassay (siehe Tab. 4-40).
Siehe beigefügte Daten-CD.

Tab. A-36: Rohdaten Human *DDX3X in-vitro* Translationsassay (siehe Tab. 4-50).
Siehe beigefügte Daten-CD.

Tab. A-37: Rohdaten Primaten *DDX3X in-vitro* Translationsassay (siehe Tab. 4-52).
Siehe beigefügte Daten-CD.

Tab. A-38: Primersequenzen zur RT-PCR Analyse von *DDX3Y* Transkripten.
Die Positionsangaben beziehen sich auf den humanen Y-chromosomalen BAC 475I1. Die Primer-Reihenfolge ist von ATG proximal zu 5´ distal. Die Primernummern beziehen sich auf die interne Labornomenklatur. F = forward, R = revers. Die RT-PCR-Serien zur Bestimmung der *DDX3Y* 5´Transkriptenden wurden bei einer optimalen Annealingtemperatur von T_A = 61˚C durchgeführt. Für Maus *Ddx3y* Transkripte wurde per Gradient das Optimum bei 57˚C festgestellt.

Primer Nr.	**Sequenz 5´-3´**	**Position in BAC 475I1**	**Länge, Schnittstelle**	**Spezies**
1261F	TCATGTGGTGGTGAAAAATGAC	55.151	22 bp	Hsap
2067F	CATATTACCGCGTAGGCTAA	55.031	20 bp	Hsap, Mmul
2405F	TAGGGAGAAGTAACGGTAGC	54.925	20 bp	Hsap, Ptro
2389F	CTACGTAATCCCGCGATCTTATG	54.885	23 bp	Hsap, Ptro, Mmul
1727F	CTTAAGGAAGAAGGGAAGTTGAG	54.699	23 bp	Hsap, Ptro, Mmul
1988F	CCACCGTGGATCGCTCAC	54.502	18 bp	Hsap
1313F	atgaac*GAGCTC*TCAAGTCTGTCGAGCCTCTG	54.362	32 bp, *SacI*	Hsap
1726F	CTGTGATGCTAAAGCCGTATGC	54.263	22 bp	Hsap, Ptro, Mmul
2278F	CTATGCCAGGGTGCGTTAAGG	53.967 + 54.066 + 54.165	21 bp	Hsap, Ptro
1330F	tacttg*GAGCTC*ACTATGCCAGGGTGCGTTAA	54.065 + 54.164	32 bp, *SacI*	Hsap
2640F	TGTTTCTTCATCACTATGCCAG	54.053 + 54.152	22 bp	Hsap
1329F	tacttg*GAGCTC*AAAACAAGGGACCTATGCCaG	53.854 + 53.955	33 bp, *SacI*	Hsap, Ptro, Mmul
1328F	tacttg*GAGCTC* CTGGCTTTGGGGACGTAGTA	53.798	32 bp, *SacI*	Hsap, Ptro
1617R	ATCTTCAGGTTAGGGAGGTG	54.491	20 bp	Hsap, Ptro
1286R	CCTTTGCTCGCTGTACTTGC	57.788	20 bp	Hsap, Mmul
1496	GATTACGCACGTTGTGCGTG	54.987	20 bp	Hsap
2420F	CATATTACCGCGTAGGTTAAGC		22 bp	Ptro

2421F	ATGACGCACGTCGCGCGTG		19 bp	Ptro
2598F	TGTGCGTTAAGGGCCTAGGT		20 bp	Ptro
2422F	TCAAGCCTGTCGAGCCTCTG		20 bp	Ptro
2419R	CCTTTGCTCGCTCTACTTGC		20 bp	Ptro
2520F	TCATGTGGGGGTGAAAAATGAC		22 bp	Mmul
2529F	TAGGCAGAAGTAATGATAGC		20 bp	Mmul
2530F	TCAAGCCTGTCGAGCTTGTG		20 bp	Mmul
2531F	CTATGCCAGTGTTCGTTAAGG		21 bp	Mmul
2599F	TGGGACCTATGCCAGTGTG		19 bp	Mmul
2601F	CTGGCTTTGGGGACTCAGTA		20 bp	Mmul
2519F	TCATGTGCTGGTGAAAAACGAC		22 bp	Cjac
2319F	ACCATTTTAAGACGGAGTCTAAG		23 bp	Cjac
2329F	ACGTCTTGCGTGCTGATGTAG		21 bp	Cjac
2388F	CTACCTGAACCCGCGATCTTAC		22 bp	Cjac
2403F	GCGGGTATAAGAAAGATATGGTG		23 bp	Cjac
2404F	GACAACTTGCGGCTAGTTGC		20 bp	Cjac
2533F	TCAAGCCCGTAAAGTATCTGG		21 bp	Cjac
2532F	CTGTGACGCTAAGGTCGTATGC		22 bp	Cjac
2534F	CTCAGGTAGTGTGTGTTAAGG		21 bp	Cjac
2600F	CGTTAAGGGCCTAGGTCCCTGG		22 bp	Cjac
2602F	CTGGCTTTGGAGACGCAAAA		20 bp	Cjac
2513R	CGTCCTTTGCTTGCTGTACTTGC		23 bp	Cjac
2568F	TCATGAGGCGGTGAAAAACGTG		22 bp	Btau
2569F	TCCGGTAGACCAACCTGTG		19 bp	Btau
2570F	TGGCCAGAAGTAACGGTAG		19 bp	Btau
2571F	TTATCGTTGACCAGTGTACG		20 bp	Btau
2589F	TAGACGCCAGCACGTGCTG		19 bp	Btau
2590F	CTGCGTGACGCGTTAACATTAG		22 bp	Btau
2591F	TTAGAATCTGCGCGGTGAGG		20 bp	Btau
2603F	TTGCCACCAGACCTAGATCC		20 bp	Btau

2567R	CCTTTGCTAGCTGTACTTCC		20 bp	Btau
2312F	GTTCCGTGAGAAGTTCTCAA		20 bp	Mmus
2313F	TGAGGAAACAGGTCTACG		18 bp	Mmus
2328F	ACTGTGTACGTAAGCTCTTGC		21 bp	Mmus
2371F	ACTCGTCAGTACGTTTGGG		19 bp	Mmus
2372F	CACGTTGTACGAATTGACGAG		21 bp	Mmus
2314F	TCTATTCTCGCACCTGAAAAGG		22 bp	Mmus
2315F	TAGAATAGTTCAGCAGAGACTAC		23 bp	Mmus
2311R	CCTTTGCTCTCTGTATTTCC		20 bp	Mmus

Tab. A-39: Primersequenzen zur RT-PCR Analyse von *DDX3X* Transkripten.

Die Positionsangaben beziehen sich auf den humanen X-chromosomalen BAC RP13-13A3. Die Primer-Reihenfolge ist von ATG proximal zu 5´ distal. Die Primernummern beziehen sich auf die interne Labornomenklatur. F = forward, R = revers. Die RT-PCR-Serien zur Bestimmung der *DDX3X* 5´Transkriptenden wurden bei einer optimalen Annealingtemperatur von T_A = 61˚C durchgeführt. Für Maus *Ddx3x* Transkripte wurde per Gradient das Optimum bei 57˚C festgestellt.

Primer Nr.	**Sequenz 5´-3´**	**Position in BAC RP13-13A3**	**Länge**	**Spezies**
2423F	CCAGAGCCGCAGTTCTCCCG	30.637	20 bp	Hsap, Ptro, Mmul, Cjac
2332F	AGCGGGTATATTAGATCCGTG	30.601	21 bp	Hsap, Mmul, Btau
910F	GCGGGAGACGCAAGTTCTCG	30.326	20 bp	Hsap
1494F	TTTCAGGCGTCTTAGCCGGT	30.301	20 bp	Hsap, Ptro
2416F	CGAAAGCTGAGTGCATAGAG	30.221	20 bp	Hsap
1493F	GTCCGGAGCACTCTATATTCA	29.959	21 bp	Hsap, Ptro
2242F	CTTTCCCCTTACTCCGCTCC	29.875	20 bp	Hsap, Ptro, Mmul
2287F	CCCGGTCACCATCCTACC	29.844	18 bp	Hsap, Ptro
2526F	ATTTCCCCCCGACTCCCGC	29.825	19 bp	Hsap, Ptro, Mmul, Cjac
2286F	AAGGTGTCCGGCCACCTG	29.785	18 bp	Hsap, Ptro
1521R	CTAGGCCAGCAAACTGCTG	30.769	19 bp	Hsap, Ptro, Mmul, Cjac, Btau, Mmus
2424F	AGCGGGTATATTAGCTCCGTG		21 bp	Ptro
2528F	GTCCCGAGCACTCTAAATTCAAGC		24 bp	Mmul
2527F	CCCGTCACCATCCTACCCTCC		21 bp	Mmul
2561F	GCTCTGGATCAGCGCCTCCGAC		22 bp	Mmul
2572F	TGTAGCCCGCATTCCTGCG		19 bp	Mmul

2560F	GAAAATGGCACAATAGGACTGTG		23 bp	Mmul
2331F	AGCGGGTATATTAGGTCCGTG		21 bp	Cjac
1971F	CACGCCACCGTTTCCTGC		18 bp	Cjac
2401F	TCCACTTCCCTCTCTTCTCTCC		22 bp	Cjac
2402F	CCCGGTGACCATCCTACC		18 bp	Cjac
2425F	CGGCTCCCCTCCATTTTCC		19 bp	Cjac
2418F	TACCGCGTAGCGAGGTGC		18 bp	Cjac
2365F	AGTTCTCCCGTGAGAGGGC		19 bp	Btau
2397F	ACGTCGTACGTGCTGACGTAG		21 bp	Btau
2398F	GTAGCTAGGCAAAAGTCGTAGC		22 bp	Btau
2366F	AGTTGCACTCAAGCGATCTAGG		22 bp	Btau
2367F	GAGACCTCCCAGGCTTGAG		19 bp	Btau
2368F	TAAGGTCTCCTAGCTGGCG		19 bp	Btau
2316F	TGAGAGGGCCTTTGCGGTA		19 bp	Mmus
2317F	TCAAAGCCGCAGTTCTCC		18 bp	Mmus
2330F	TCGGTATATAAGGTCGGTGG		20 bp	Mmus
2318F	CTTAGGGCCTTAAGGCTAGAG		21 bp	Mmus

Tab. A-40: Primersequenzen zur cDNA-Kontrolle (Beta-Actin) per RT-PCR-Analyse.

Primer Nr.	Name	Sequenz 5´-3´	Länge	Spezies
518F	Beta-Actin	ATCCTGCGTCTGGACCT	18 bp	Hsap, Ptro, Mmul
519R	Beta-Actin	GCTGATCCACATCTGCTG	18 bp	Hsap, Ptro, Mmul, Cjac
2559F	Beta-Actin	ATCCTGCGTCTGGACCTG	18 bp	Cjac
2399F	Beta-Actin	GAAGATGACCCAGATCATGTTCG	23 bp	Btau
2400R	Beta-Actin	CTGCATCCTGTCCTCGATGC	20 bp	Btau

Tab. A-41: Primersequenzen zur Insertkontrolle per PCR-Analyse und Sequenzierung.

Primer Nr.	Name	Sequenz 5´-3´	Länge	Spezies
968F	pGEM-T Easy	TAATACGACTCACTATAGGG	20 bp	T7 für PCR und Seq.
1621R	pGEM-T Easy	ATACGATTTAGGTGACACTATAGAA	25 bp	SP6 für PCR und Seq.
1316F	pGL4.10, pGL4.13	CTAGCAAAATAGGCTGTCCC	20 bp	RVprimer3 für PCR und Seq.
1830R	pGL4.10, pGL4.13	GTCTTCGAGTGGGTAGAATG	20 bp	pGL4rev. für PCR und Seq.

Tab. A-42: Primersequenzen zur Klonierung der Human *DDX3X* Promoterregion und der *MSY2*/*MSY2-X* Sequenzen.

Nukleotidaustausche für site-directed Mutagenesis sind rot markiert.

Primer Nr.	Sequenz 5´-3´	Konstruktbezeichnung	Schnittstelle
1887F	tacttgc*GAGCTC*gATGACTGGCGAGAGCTACTG	*DDX3X* Konstrukt I	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1312F	tacttg*GAGCTC*gCTCATGGAACTTGAGTTGTG	*DDX3X* Konstrukt II	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1897F	tacttg*GAGCTC*gTTGCTTTCGTTTGTTGACAGATC	*DDX3X* Konstrukt III	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1995F	tacttg*GAGCTC*gCACAGGGTTATGGAATGTAGC	*DDX3X* Konstrukt IV	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1994F	tacttg*GAGCTC*gGCGGTCATAAGAAACTACGG	*DDX3X* Konstrukt V	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1305F	tactag*GAGCTC*gGCGGTCATAAGAAACTACGG	*DDX3X* Konstrukt VI	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1388F	tactag*GAGCTC*CTGTCTCATTGTTCCCTCAC	*DDX3X* Konstrukt VII	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1389F	tactag*GAGCTC*CGGTCACCATCCTACCC	*DDX3X* Konstrukt VIII	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1390F	tactag*GAGCTC*TCTCCAAGCCCGGGAACTG	*DDX3X* Konstrukt IX	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1391F	tactag*GAGCTC*CTCCGAGGCTGAGACTAGG	*DDX3X* Konstrukt X	*SacI*
1829R	tccta*GCTAGC*tagtAGCCCGAGCGCATTTTCCA		*NheI*
1991F	ccc*AAGCTT*gggTCTCCAAGCCCGGGAACTG	*DDX3X* Insert a	*HindIII*
1992R	ccc*AAGCTT*gggCCTAGTCTCAGCCTCGGAG		
1991F	ccc*AAGCTT*gggTCTCCAAGCCCGGGAACTG	*DDX3X* Insert b	*HindIII*
1993R	ccc*AAGCTT*gggCTCCCTCTCACACCGCAATTCC		
1860F	ccc*AAGCTT*gggGCGGTCATAAGAAACTACGG	*DDX3X* Insert c	*HindIII*
1861R	ccc*AAGCTT*gggGAGTAAGGGGAAAGgaggag		
1989F	ccc*AAGCTT*gggGCTACAAACATTCCCgACAC	*DDX3X* Insert d	*HindIII*
1861R	ccc*AAGCTT*gggGAGTAAGGGGAAAGgaggag		
1990F	ccc*AAGCTT*gggCTGTCTCATTGTTCCCTCAC	*DDX3X* Insert e	*HindIII*
1861R	ccc*AAGCTT*gggGAGTAAGGGGAAAGgaggag		
2299F	ccc*AAGCTT*gggAAGGTGTCCGGCCACCTG	*DDX3X* Insert f	*HindIII*
2300R	ccc*AAGCTT*gggCTTGAATATAGAGTGCTCC		
1882F	ccc*AAGCTT*gggCTAGTTCATTATCATCTAAGCC	*DDX3X* Insert g	*HindIII*
1898R	ccc*AAGCTT*gggAGGGTAGGATGGTGACCG		
1896F	ccc*AAGCTT*gggTTGCTTTCGTTTGTTGACAGATC	*DDX3X* Insert h	*HindIII*
1898R	ccc*AAGCTT*gggAGGGTAGGATGGTGACCG		
1996F	ccc*AAGCTT*gggCACAGGGTTATGGAATGTAGC	*DDX3X* Insert i	*HindIII*
1898R	ccc*AAGCTT*gggAGGGTAGGATGGTGACCG		
1882F	ccc*AAGCTT*gggCTAGTTCATTATCATCTAAGCC	*DDX3X* Insert j	*HindIII*
1883R	cccAAGCTTgggATCTGTCAACAAACGAAAGC		

1882F 1884R	ccc*AAGCTT*gggCTAGTTCATTATCATCTAAGCC ccc*AAGCTT*gggAGGAGTCTGCTACATTCCA	*DDX3X* Insert k	*HindIII*
1896F 2059R	ccc*AAGCTT*gggTTGCTTTCGTTTGTTGACAGATC ccc*AAGCTT*gggTTGGGTCTCAATTTGCTCATC	*DDX3X* Insert l	*HindIII*
2408F 2406R	ccc*AAGCTT*gggCTGGCTTTGGGGACGTAGTA ccc*AAGCTT*gggCACGAAACAATGAAACAATTATATAC	*DDX3Y MSY2* Insert m	*HindIII*
2408F 2407R	ccc*AAGCTT*gggCTGGCTTTGGGGACGTAGTA ccc*AAGCTT*gggGATGAAGAAACAATGAAACAATTAC	*DDX3Y MSY2* Insert n, Insert o	*HindIII*
2408F 2411R	ccc*AAGCTT*ggg CTGGCTTTGGGGACGTAGTA ccc*AAGCTT*gggCTGCGCTTTATGAGCATACG	*DDX3Y MSY2* Insert p	*HindIII*
2409F 2411R	ccc*AAGCTT*ggg AAAACAAGGGACCTATGCCAG ccc*AAGCTT*gggCTGCGCTTTATGAGCATACG	*DDX3Y MSY2* Insert q, Insert r	*HindIII*
2410F 2411R	ccc*AAGCTT*gggACTATGCCAGGGTGCGTTAAG ccc*AAGCTT*gggCTGCGCTTTATGAGCATACG	*DDX3Y MSY2* Insert s, Insert t, Insert u	*HindIII*
2352F	GTGTGTTATCACCAGGGCCATTCAGTATATAATTG	*DDX3Y MSY2*	Mutation VDR
2353R	CAATTATATACTGAATGGCCCTGGTGATAACACAC	*DDX3Y MSY2*	Mutation VDR
2354F	GGTGTGTTATAATCAACGCCATTCAGTATG	*DDX3Y MSY2*	Mutation CREB1
2355R	CATACTGAATGGCGTTGATTATAACACACC	*DDX3Y MSY2*	Mutation CREB1
2294F 2295R	ccc*AAGCTT*gggGAAACTAGGGGCAAATGTAGC ccc*AAGCTT*gggGAGGGAACAATGAGACAGACA	*DDX3X MSY2-X* Insert v	*HindIII*
2294F 2296R	ccc*AAGCTT*gggGAAACTAGGGGCAAATGTAGC ccc*AAGCTT*gggGCCTAGAAACACACAATGG	*DDX3X MSY2-X* Insert w	*HindIII*
2348F	CGCGTGTCTGTCTCACTGTTCCCTC	*DDX3X MSY2-X*	Mutation Sox5
2349R	GAGGGAACAGTGAGACAGACACGCG	*DDX3X MSY2-X*	Mutation Sox5
2350F	GGATCAGCGCGTGCCTGTCTCATTGTTC	*DDX3X MSY2-X*	Mutation SMAD3
2351R	GAACAATGAGACAGGCACGCGCTGATCC	*DDX3X MSY2-X*	Mutation SMAD3

Tab. A-43: Primersequenzen zur Klonierung der Human und Primaten *DDX3X* 5´UTR-Sequenzen.
Primersequenzen und Positionen im BAC RP13-13A3, unmodifizierte und modifizierte Primer. Ø = keine Modifikation am Primer. Enzymschnittstellen sind kursiv gegeben, Linker zwischen Schnittstellen und komplementärer Sequenzen sind mit Kleinbuchstaben, T7-RNA Polymerase Promotersequenz ist unterstrichen, komplementäre Sequenzen zu *DDX3X* und *Luc*2 sind fett markiert, Kozak-Sequenz ist umrahmt, künstlicher Poly(A)-Schwanz ist türkis, Stopcodone sind grün unterlegt. Nukleotidaustausche für site-directed Mutagenesis sind rot markiert.

2509F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**ATTTCCCCCCGACTCCCG** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H1, H2	*Hind*III + T7-Promoter; *Kpn*I
2244F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**CGAAGAGGCCTAGGTTAACA** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H3	*Hind*III + T7-Promoter; *Kpn*I
2245F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**CCAGAGCCGCAGTTCTCCCG** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H4	*Hind*III + T7-Promoter; *Kpn*I
2654F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**GAAAGCTGAGTGCATAGAGTTGG** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H5	*Hind*III + T7-Promoter; *Kpn*I

2430F 2246R	cccAAGCTTgggTAATACGACTCACTATAGGG**GTCCGGAGCACTCTATATTCA** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H6	*Hind*III + T7-Promoter; *Kpn*I
2509F 2460R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**ATTTCCCCCCGACTCCCG** gg*GGTACC*cc**TCCCGCTTTCCACCGGCTAAGACG**	*DDX3X* H7	*Hind*III + T7-Promoter; *Kpn*I
2509F 2648R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**ATTTCCCCCCGACTCCCG** gg*GGTACC*cc**CTTCACTCTTCTTCCTCTCTCAG**	*DDX3X* H8	*Hind*III + T7-Promoter; *Kpn*I
2243F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**CTTTCCCCTTACTCCGCTCC** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* H9, H10	*Hind*III + T7-Promoter; *Kpn*I
2243F 2460R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**CTTTCCCCTTACTCCGCTCC** gg*GGTACC*cc**TCCCGCTTTCCACCGGCTAAGACG**	*DDX3X* H11	*Hind*III + T7-Promoter; *Kpn*I
2243F 2648R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**CTTTCCCCTTACTCCGCTCC** gg*GGTACC*cc**CTTCACTCTTCTTCCTCTCTCAG**	*DDX3X* H12	*Hind*III + T7-Promoter; *Kpn*I
2458F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**GAGGCCTAGGTTAACATTTTCAGG** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* P3	*Hind*III + T7-Promoter; *Kpn*I
2430F 2246R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG**GTCCGGAGCACTCTATATTCA** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* P6	*Hind*III + T7-Promoter; *Kpn*I
2652F 2246R	cccAAGCTTgggTAATACGACTCACTATAGGG GTCCCGAGCACTCTAAATTCA gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* M3	*Hind*III + T7-Promoter; *Kpn*I
2653F 2246R	cccAAGCTTgggTAATACGACTCACTATAGGG GAGGCCTAGGTTAACATTTCCAGG gg*GGTACC*cc**AAGAGTACCGAGAACTCGGAGTC**	*DDX3X* M6	*Hind*III + T7-Promoter; *Kpn*I
2428F 2429R	ccc*AAGCTT*gggTAATACGACTCACTATAGGG **CCCGGTGACCATCCTACC** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGTGTC**	*DDX3X* C1	*Hind*III + T7-Promoter; *Kpn*I
2459F 2429R	cccAAGCTTgggTAATACGACTCACTATAGGG**GAGGCTTAGATTAACATTTTCAGG** gg*GGTACC*cc**AAGAGTACCGAGAACTCGGTGTC**	*DDX3X* C3	*Hind*III + T7-Promoter; *Kpn*I
2510F	GTTTCCTGCTTCACAAATAGGCCACCGCACGCGAC	*DDX3X*	Mutation ATG uORF1; Hsap, Ptro
2511R	GTCGCGTGCGGTGGCCTATTTGTGAAGCAGGAAAC	*DDX3X*	Mutation ATG uORF1; Hsap, Ptro
2247F	gg*GGTACC*ccCAGGG**ATG**GAAGATGCCAAAAACATTAAG	*Luc*2	*Kpn*I, Kozak-Seq.
2248R	ccg*GAATTC*cgg**TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**AGAATTATTACACGGCG ATCTTGCCGC	*Luc*2	*Eco*RI, Poly(A)$_{30}$
2412F	CTGCAAGGCGATTAAGTTGG	pUC18	für PCR und Seq.
2413R	GAGCGGATAACAATTTCACAC	pUC18	für PCR und Seq.
2526F	ATTTCCCCCCGACTCCCGC	*DDX3X*	Pos. 29.825
2242F	CTTTCCCCTTACTCCGCTCC	*DDX3X*	pos. 29.875
2423F	CCAGAGCCGCAGTTCTCCCG	*DDX3X*	Pos. 30.637

11. Veröffentlichungen

Teilergebnisse der vorliegenden Dissertation wurden in folgenden Beiträgen vorab veröffentlicht:

Publikation

aus Kapitel 4.1. und Kapitel 4.4.:

Rauschendorf M-A, Zimmer J, Hanstein R, Dickemann C, Vogt PH (2011)
Complex transcriptional control of the AZFa gene *DDX3Y* in human testis
International Journal of Andrology. 34(1):84-96.

Tagungsbeiträge

aus Kapitel 4.1. und Kapitel 4.4.:

Rauschendorf M-A, Zimmer J, Hanstein R, Vogt PH
5´UTR extensions of *DDX3Y* testis transcripts points to evolution of a primate-specific control mechanism for expression of *DDX3Y* in male germ cells
5th International Workshop on Molecular Andrology
Gießen 08.05. – 10.05.2009

Rauschendorf M-A, Zimmer J, Hanstein R, Vogt PH
5´UTR extensions of *DDX3Y* testis transcripts points to evolution of a primate-specific control mechanism for expression of *DDX3Y* in male germ cells
20. Jahrestagung der Deutschen Gesellschaft für Humangenetik
Aachen 01.04. – 03.04.2009

aus Kapitel 4.6. und Kapitel 4.9.

Rauschendorf M-A, Zimmer J, Hanstein R, Vogt PH
Divergent transcriptional control of the AZFa gene *DBY (DDX3Y)* and its X-homologue *DBX (DDX3X)* by allel-specific promoter structure
19. Jahrestagung der Deutschen Gesellschaft für Humangenetik
Hannover 08.04. – 10.04.2008

12. Eidesstattliche Erklärung

Erklärung:

Ich versichere hiermit gemäß §8 (3) b) und c) der Promotionsordnung, daß ich die vorgelegte Doktorarbeit

„Molekulare Struktur- und Funktionsanalyse der Transkriptionskontrolle der Gene *DDX3X* und *DDX3Y* in der männlichen Keimbahn“

unter der Leitung von Herrn Prof. Dr. Hans-Peter Vogt (Universitätsfrauenklinik Heidelberg) selbstständig und ohne unerlaubte Hilfe angefertigt und mich keiner anderen als der von mir ausdrücklich angegebenen Quellen und Hilfen bedient habe.

Heidelberg, den 05. Juli 2011

_____

Marc-Alexander Rauschendorf

13. Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. HP. Vogt für die Betreuung und seine Bereitschaft die Arbeit als Erstgutachter vor der Fakultät für Biowissenschaften zu vertreten. Die fachliche Unterstützung, die anregenden Diskussionen und die produktive Zusammenarbeit waren eine große Hilfe bei der Anfertigung dieser Arbeit.

Herrn Prof. Dr. H. Steinbeisser danke ich sehr herzlich für die Übernahme des Zweitgutachtens.

Ich danke Herrn Dr. L. Jaroszynski für die vielen hilfreichen wissenschaftlichen Diskussionen.

Besonderer Dank gilt Frau J. Zimmer für ihre vielen technischen und methodischen Hilfestellungen.

Mein Dank gilt auch Frau U. Bender für ihre vielen organisatorischen Hilfestellungen.

Herzlich bedanken möchte ich mich bei Frau B. Güler und Frau Dr. C. Falcao für ihre Hilfsbereitschaft, sowie für das angenehme Arbeitsklima.

Bedanken möchte ich mich bei Frau N. Fröhlich und Herrn M. Vogel für ihre Unterstützung.

Ich danke allen ehemaligen und aktuellen Mitarbeitern der Arbeitsgruppe für die freundliche Unterstützung, sowie das angenehme Arbeitsklima.

Mein innigster Dank gilt meinen Eltern und meiner Familie für ihre unermüdliche Unterstützung und ständige Hilfsbereitschaft. Ohne ihr Verständnis in schwierigen Phasen, wäre diese Arbeit nicht möglich gewesen.